SAMPLING THEORY IN SIGNAL AND IMAGE PROCESSING
Vol. 7, No. 2, May 2008, pp. 141-171
ISSN: 1530-6429
© 2008 SAMPLING PUBLISHING

# Halton Sampling for Image Registration Based on Mutual Information

Philippe Thévenaz
École polytechnique fédérale de Lausanne (EPFL)
Biomedical Imaging Group, Station 17, CH–1015 Lausanne VD, Switzerland
philippe.thevenaz@epfl.ch

Michel Bierlaire
École polytechnique fédérale de Lausanne (EPFL)
Transport and Mobility Laboratory, Station 18, CH–1015 Lausanne VD, Switzerland
michel.bierlaire@epfl.ch

Michael Unser
École polytechnique fédérale de Lausanne (EPFL)
Biomedical Imaging Group, Station 17, CH–1015 Lausanne VD, Switzerland
michael.unser@epfl.ch

## Abstract

Mutual information is a widely used similarity measure for aligning multimodal medical images. At its core it relies on the computation of a discrete joint histogram, which itself requires image samples for its estimation. In this paper we study the influence of the sampling process. We show that quasi-random sampling based on Halton sequences outperforms methods based on regular sampling or on random sampling. Our results suggest that sampling itself—and not interpolation, as was previously believed—is the source of two major problems associated with mutual information: the grid effect, whereby grid-aligning transformations are favored, and the overlap problem, whereby the similarity measure exhibits discontinuities. Both defects tend to impede the accuracy of registration; they also result in reduced robustness because of the presence of local optima. By estimating the joint histogram by quasi-random sampling, we solve both issues at the same time.

*Keywords and phrases*: Image registration, mutual information, multimodal medical images, Halton sampling

*2000 AMS Mathematics Subject Classification*— 94A17, 92C50

## 1 Introduction

The formulation of a registration task usually requires three basic ingredients: a spatial transformation model to define the space of possible solutions, an objective similarity measure to estimate the quality of fit of each potential solution, and an optimization algorithm to identify which of all solutions within this space results in the best quality of fit [2, 8, 24, 26]. The type of registration task imposes specific design requirements for each of these three ingredients.

The focus of this paper is on the similarity measure, for which many forms have been proposed, for example, based on moments [12], on Fourier [35], on optical flow [22], on landmarks [1], on the mean-square error [37], or on a combination of mean squares and landmarks [33], among others. In the specific context of multimodal registration, this measure ought to be able to cope with images for which the mapping between intensities is not necessarily one-to-one; in a variety of applications [6, 17], good success has been reported when using the mutual information, first proposed for signal alignment in [25]. The justification for this choice, and the comparison of the merits of mutual information to those of several other similarity measures, can be found elsewhere [28, 31].

The mutual information between two images is obtained by subtracting their entropy from their joint entropy. This results in a measure of the degree of the mutual dependence of the gray levels of two images, or, put differently, of the degree to which a given gray level of one image can predict gray levels in the other. It is therefore an attractive criterion to register images or volumes, assuming that their correct alignment coincides with that for which the gray levels of one image collectively best explain those of the other. Mutual information is particularly useful when the images are acquired through different modalities because it does not depend on the specific mapping between their gray levels. It also offers some degree of robustness to features that would be discernible in one modality but absent from the other. It is also very well suited to registration problems where illumination is uniform, such as many of the problems that are prevalent in medical imaging where it is desirable to bring into registration data obtained by several types of scanner. However, as it relies on an analysis of the gray levels that is global, it is not well suited to tasks where the statistical aspects of the gray levels vary within the image, such as may typically be the case with natural images which introduce shadows, vignetting, and overall gradients due to directional illumination.

At its heart, the mutual information relies on the computation of a joint histogram between discrete intensities of the images or volumes to register. Its computation can be performed in a variety of ways; in particular, it depends strongly on details of the method of computing the joint histogram. An innovation of this paper is to propose that the samples that are used to build the joint histogram are taken at truly arbitrary coordinates. Rather than applying a random perturbation to integer coordinates, as was done in [21], we let the location of the samples follow a random pattern that is not related to integer coordinates anymore. This offers us more freedom to design the registration method by giving full control over the number of samples and by allowing us to carefully choose the random sampling pattern itself. The conclusions regarding the benefits of our approach are applicable to several other contexts as well, such as estimations of the joint histogram based on the traditional partial-volume method or on its generalized version [4, 5], or estimation of the joint probability density function based on Gaussian Parzen windows.

Here, we investigate novel issues that take relevance only in the context of random sampling, but not in that of regular sampling. We propose two key contributions and perform their systematic evaluation. These contributions are:

- *Halton sampling.* The selection of samples follows a quasi-random Halton sequence in this paper. This sequence enjoys a property called low-discrepancy [7], which ensures that data are represented by their random samples in a way that is much more faithful than if the probability density function of the samples would be uniformly distributed. While Halton sequences are well known [44], to the best of our knowledge it is the first time that they are applied in a registration context.

- *Overlap independence.* The use of a Halton sequence allows us to completely untie the link between image-defining pixels and histogram-defining samples since the randomness applies globally, as opposed to having local randomness attached to a pixel—or no randomness at all in the case of regular sampling. This allows for an arbitrary choice of the number of samples that does not depend in any way on the number of pixels. This is an attractive feature that will be exploited as follows:

    1. *Very small or very large data.* Multiresolution is a seductive approach for registering data. For methods that rely on a joint histogram, however, it is necessary to pay attention to the dwindling statistical power at coarse resolution, for example by introducing prior joint probabilities or by reducing the number of gray levels. Unfortunately, when the original data are intrinsically small, priors may not exist or may be difficult to come by; meanwhile, maintaining the statistical power of the joint histogram may require an excessive coarsening of the quantization of the gray levels.

       In the present paper, we could compensate a low statistical power by arbitrarily choosing a sufficient number of samples, without being constrained by the size of the data or by the number of gray levels. Conversely, when the data are large, we could conveniently choose to pick less samples than there are pixels without concern for the actual downsampling pattern. In practice, however, for irregular sampling we observe the disappearance of those local extrema of the registration criterion that were associated with regular sampling; the remaining local extrema are benign and our method is so robust that there is no need for multiresolution altogether.

2. *Restored continuity.* In many implementations where a binary mask is used to define the domain of overlap of the images to register, a discrete change of overlap would result in a discontinuous similarity measure, which is difficult to optimize. In the present paper, we sever the dependence between the number of samples and the overlap; in this way, we take advantage of the new freedom to escape the discontinuities associated with the overlap problem altogether.

To perform our experimental evaluations, we have computed the joint histogram thanks to an estimate based on Parzen windows made of splines. We have used cubic-spline interpolation—a high-quality interpolation method—to apply the geometric transformations involved in the alignment of an image with respect to the other. Our experiments allow us to clearly demonstrate and quantify the advantages of using random sampling in general, of using Halton sampling in particular, and of uncoupling the number of samples from the number of pixels.

This paper is organized as follows: in Section 2 we introduce notations and definitions. In Section 3 we present two published methods for computing a joint histogram. We explain how to generate a Halton sequence in Section 4. In Section 5 we first give a methodic illustration of the problems associated with regular sampling, and then solve them; we finally validate our approach according to a methodology that involves the blind-folded comparison to a gold standard. We conclude in Section 6.

## 2 Notations and Definitions

### 2.1 Spatial Transformation Model

To align the source (or test, or floating) image $g$ to the target (or reference) image $f$, we apply to $g$ a geometric transformation $T$ that depends on a series of parameters that we collect in a vector $\mathbf{a}$. This spatial transformation model is such that a coordinate $\mathbf{x}$ of the target image $f$ is made to correspond to a coordinate $\mathbf{y} = T(\mathbf{x};\mathbf{a})$ of the source image $g$; the best transformation that maps $g$ to $T\{g\}$—respectively, that maps the sample $g(\mathbf{x})$ to $g(T(\mathbf{x},\mathbf{a}))$—is obtained for the vector of parameters $\hat{\mathbf{a}}$ for which the similarity measure reaches an optimum. The theoretical underpinnings of this paper do not depend on any specific type of transformation model, even though the experiments are all performed within a rigid-body transformation framework.

## 2.2 Mutual Information

The mutual information $I$ of images $f$ and $g$ is defined by

$$I\{f,g;\mathbb{F},\mathbb{G}\} = \sum_{\phi\in\mathbb{F}}\sum_{\gamma\in\mathbb{G}} h[\phi,\gamma]\,\log_2\frac{N\,h[\phi,\gamma]}{h_f[\phi]\,h_g[\gamma]}, \tag{1}$$

where $h[\cdot,\cdot]$ is the discrete joint histogram of the two images, where $N = \sum_{\phi\in\mathbb{F}}\sum_{\gamma\in\mathbb{G}} h[\phi,\gamma]$ is the global weight of $h$, where $h_f[\cdot] = \sum_{\gamma\in\mathbb{G}} h[\cdot,\gamma]$ and $h_g[\cdot] = \sum_{\phi\in\mathbb{F}} h[\phi,\cdot]$ are the marginal histograms of $h$, and where $\mathbb{F}$ and $\mathbb{G}$ are discrete sets of gray levels associated to $f$ and $g$, respectively. As it is customary to measure the mutual information in bit units, the base of the logarithm is chosen to be 2.

The focus of this paper is on the computation of the joint histogram $h$ which is also at the heart of several other similarity measures, such as the joint entropy $H\{f,g\} = -\sum_{\phi\in\mathbb{F}}\sum_{\gamma\in\mathbb{G}}\frac{h[\phi,\gamma]}{N}\,\log_2(\frac{h[\phi,\gamma]}{N})$, or the normalized mutual information $Y = 1 + I/H$.

We want to mention here two key points of (1). First, we observe that the notion of mutual dependence between $f$ and $g$ is explicitly captured by the ratio $h/(h_f\,h_g)$. For example, should the two images be independent, this ratio would be $1/N$ and the contribution of the pair of gray levels $\{\phi,\gamma\}$ to the mutual information would vanish. Second, it is essential to keep in mind at all times that $\mathbb{F}$ and $\mathbb{G}$ are discrete. This is inherent to the definition of mutual information; moreover, it would be wrong to assume that, by varying these greyscale sets, one would simply perform various estimates of a hypothetical gold-standard Mutual Information. There is no such thing. In particular, (1) has the unusual property that it does not converge to a finite value when the quantization of $\mathbb{F}$ and $\mathbb{G}$ is made to be finer and finer.

## 2.3 Domain of Overlap

For every similarity measure it is customary to introduce a binary mask $F$ that will indicate by $F(\mathbf{x}) = 1$ which coordinates $\mathbf{x}$ of $f$ are relevant, and by $F(\mathbf{x}) = 0$ which are not. Typically, this mask is used to reject part of the background—if so desired—or to reject parts of the data where it is expected beforehand that matching would be meaningless (*e.g.*, over tissue that would be present in $f$ but absent from $g$ because of some surgical procedure). Additionally, a mask $G$ indicates the relevance of the coordinates of $g$. Then, the similarity measure is computed only over those coordinates that are relevant with respect to both $f$ and $T\{g\}$. The set of these coordinates is called the domain of overlap $\mathbb{D} = \{\mathbf{x} : F(\mathbf{x})\,G(T(\mathbf{x})) = 1\}$.

# 3 Previous Work

On one hand, the determination of a joint histogram is a central aspect of every registration algorithm based on mutual information or on its various versions, whether directly based on information-theoretic grounds [23], or indirectly [16]. On the other hand, these algorithms must also perform an optimization task, which is made easier and much more efficient by the availability of the gradient $\partial I/\partial \mathbf{a}$ of the mutual information $I$ with respect to the deformation parameters $\mathbf{a}$ that control the geometric transformation $T$ used to align $g$ to $f$. Therefore, a differentiable $I$ is a desirable asset, and so is a differentiable $h$. This is the main reason why the trivial binning approach does not qualify when it comes to determining a joint histogram, since its gradient with respect to the deformation parameters is ill-defined.

We now revisit two methods of the literature that overcome this limitation. These two methods are very different: the first one may sometimes populate the entry $h[\phi, \gamma]$ of the joint histogram even if no pixel of $f$ ever takes the gray value $\phi$ (same for $g$ and $\gamma$), while the second method avoids this state of affairs. The first method may simultaneously update up to four times more entries of $h$ than the second one; it also considers many more image pixels and volume voxels (four times in 2-D, and eight times in 3-D) to perform one histogram update. A definite interpolation model is a necessary ingredient of the first method, while the second method requires none. Despite these deep differences, we shall see in Section 5 that the two methods of the present section suffer from very similar artifacts under regular-sampling conditions, while these artifacts disappear for quasi-random sampling. Therefore, more than anything else, proper sampling appears to be a central aspect of the determination of a joint histogram.

## 3.1 Spline Parzen (SP)

In general, the nonparametric estimation of a density function requires a discrete set of $K$ samples taken at abscissa $x_k$. In one dimension the Parzen estimate $\tilde{p}_\xi$ of a density function $p$ is given by

$$\tilde{p}_\xi(x) = \frac{1}{K} \sum_{k=0}^{K-1} \frac{1}{\xi} \, \varphi(\frac{x - x_k}{\xi}),$$

where $\varphi(\cdot/\xi)/\xi$ is the kernel of the estimate, also called the Parzen window, and where $\xi$ is a scale factor that ought to depend on $K$. This dependence should obey certain technical conditions for $\tilde{p}_\xi$ to be asymptotically unbiased and consistent; in addition, all kernels are not admissible. A full discussion of the technical requirements is available in [13]. In $q$ dimensions it is usual to consider that kernels are given by a Cartesian tensor product such that

$$\tilde{p}_\xi(\mathbf{x}) = \frac{1}{K} \sum_{k=0}^{K-1} \prod_{i=1}^{q} \frac{1}{\xi_i}\, \varphi_i(\frac{[\mathbf{x} - \mathbf{x}_k]_i}{\xi_i}), \tag{2}$$

where the operator $[\,\cdot\,]_i$ must be understood here as returning the $i$-th component of its vector argument.

In [39], we have proposed a method that is based on Parzen windows with B-splines in the role of the weight functions. Rewriting (2) with the symmetric quadratic B-spline $\beta^2(\cdot) = \varphi(\cdot/\xi)/\xi$ in the role of the kernel of the estimate, and multiplying $\tilde{p}_\xi$ by the number of samples $K$ to consider a histogram instead of a probability density, we obtain a first form of the Parzen estimate of the joint histogram

$$\bar{h}(s,t) = \sum_{\mathbf{x}_k \in \mathbb{D}} \beta^2(f(\mathbf{x}_k) - s)\, \beta^2(g(T(\mathbf{x}_k)) - t).$$

In this expression the joint intensities $(s,t)$ may take values in a continuum.

Since the computation of (1) requires sets of intensities $\mathbb{F}$ and $\mathbb{G}$ that must be discrete, we need to discretize the continuous Parzen estimate $\bar{h}$. Rather than simply evaluating $\bar{h}$ for a discrete set of arguments $\{(\phi,\gamma)\} \in \mathbb{F} \times \mathbb{G}$, we prefer to average all values inside a bin of unit width centered on $\phi$ and $\gamma$. We write

$$h(\phi,\gamma) = \int_{\phi-\frac{1}{2}}^{\phi+\frac{1}{2}} \int_{\gamma-\frac{1}{2}}^{\gamma+\frac{1}{2}} \bar{h}(s,t)\,\mathrm{d}s\,\mathrm{d}t$$

which, due to the property $\int_{x-\frac{1}{2}}^{x+\frac{1}{2}} \beta^n(t)\,\mathrm{d}t = \beta^{n+1}(x)$, we can also rewrite as

$$h[\phi,\gamma] = \sum_{\mathbf{x}_k \in \mathbb{D}} \beta^3(f(\mathbf{x}_k) - \phi)\, \beta^3(g(T(\mathbf{x}_k)) - \gamma). \tag{3}$$

This allows us to produce a simple expression of the gradient $\partial I/\partial \mathbf{a}$. One component of the corresponding gradient is then expressed as

$$\begin{aligned} \frac{\partial h}{\partial a} &= \sum_{\mathbf{x}_k \in \mathbb{D}} \beta^3(f(\mathbf{x}_k) - \phi) \left(\frac{\partial}{\partial a} g(T(\mathbf{x}_k;\mathbf{a}))\right) \\ &\quad \cdot \underbrace{\left(\beta^2(g(T(\mathbf{x}_k)) - \gamma + \frac{1}{2}) - \beta^2(g(T(\mathbf{x}_k)) - \gamma - \frac{1}{2})\right)}_{\frac{\mathrm{d}\beta^3(x)}{\mathrm{d}x}\big|_{x = g(T(\mathbf{x}_k)) - \gamma}}. \end{aligned} \tag{4}$$

We observe that interpolation of $g$ is required since terms of the form $g(T(\mathbf{x}))$ do appear in (3) and (4). To avoid having to interpolate $f$, the condition $\mathbb{D} \subseteq \mathbb{Z}^q$ was assumed in [39] but will be relaxed in this paper. Finally, the mutual information is computed according to (1) with explicit $\mathbb{F}$ and $\mathbb{G}$. In the present paper we call this approach “Spline Parzen” (SP).

## 3.2 Image Model

The registration of $g$ with respect to $f$ involves a transformation $T$ such that a coordinate $\mathbf{x}$ in $f$ is put into correspondence with a coordinate $T(\mathbf{x})$ in $g$. Since $f$ and $g$ are sampled, this correspondence takes an unambiguous meaning only for those pairs of coordinates $\mathbf{x}$ and $T(\mathbf{x})$ that coincide with the coordinates of existing samples $\{\mathbf{x}_i, f(\mathbf{x}_i)\}$ and $\{\mathbf{x}_j, g(\mathbf{x}_j)\}$, with $\mathbf{x}_j = T(\mathbf{x}_i)$. In principle, a fair similarity measure should take only such coincident pairs into account. Unfortunately, there are very few coincident pairs—often none—because it is almost always the case that the transformation $T$ is such that $f(\mathbf{x}_i)$ and $g(T(\mathbf{x}_i))$ cannot be both well-defined by the available data samples.

A better solution is to introduce an interpolation model to allow the evaluation of $g$ over a continuous range of arguments. This is the approach followed by SP, with a firm theoretical grip on the resulting quality of the interpolation [38]. In our case a single interpolation model will allow us to handle at the same time the determination of $f(\mathbf{x})$ or $g(T(\mathbf{x}))$ for non-integer $\mathbf{x}$ or $T(\mathbf{x})$, and the determination of the spatial gradient $\boldsymbol{\nabla} g(T(\mathbf{x}))$. In 1-D this model is given by

$$f(x) = \sum_{k \in \mathbb{Z}} c_k \, \beta^n(x - k),$$

where the B-spline of degree $n$ is computed as a weighted sum of polynomial simple elements $\varsigma^n(x) = \frac{1}{2\,n!} \operatorname{sgn}(x)\, x^n$; more precisely, it is given by $\beta^n(x) = \sum_{m=0}^{n+1} (-1)^m \binom{n+1}{m} \varsigma^n(x + \frac{n+1}{2} - m)$. Because B-splines have a low error of approximation, the use of this model ensures high fidelity to the data, as long as the coefficients $c_k$ are computed properly [38, 40]. Moreover, the spatial gradient is easy to obtain. It can be expressed as

$$\frac{\partial f}{\partial x} = \sum_{k \in \mathbb{Z}} c_k \, \dot{\beta}^n(x - k) = \sum_{k \in \mathbb{Z}} (c_k - c_{k-1}) \, \beta^{n-1}(x + \frac{1}{2} - k),$$

where $\dot{\beta}^n(x) = \sum_{m=0}^{n+1} (-1)^m \binom{n+1}{m} \varsigma^{n-1}(x + \frac{n+1}{2} - m)$.

The expressions above are valid for all $x \in \mathbb{R}$. The extension to $\mathbb{R}^q$ is obtained by computing $\sum c_{\mathbf{k}} \, \beta^n(\mathbf{x} - \mathbf{k})$, where $\beta^n(\mathbf{x}) = \prod_{i=1}^{q} \beta^n([\mathbf{x}]_i)$, $\mathbf{x} \in \mathbb{R}^q$, is the Cartesian tensor product of unidimensional B-splines of degree $n$. Therefore, these expressions can be applied directly to interpolate $f$ or $g$ or $\boldsymbol{\nabla} g$ over the whole domain $\mathbb{D}$. We have used cubic splines throughout this paper, except in the cases that involve partial volume, for which no explicit interpolation takes place.

## 3.3 Partial Volume

In the approach called "Partial Volume" (PV), an explicit joint histogram is built as a sum of independent contributions, while the need to build an explicit transformed image is avoided—this is peculiar to the PV method. To understand this approach, suppose that a coordinate $\mathbf{k}$ of the reference image $f$ is aligned with a coordinate $T(\mathbf{k})$ of the test image $g$. If one would produce $g(T(\mathbf{k}))$ by linear interpolation, one would need to compute a weighted sum of the intensities $g([T(\mathbf{k})] + \Delta\mathbf{k})$. Here, $[\,\cdot\,]$ indicates rounding to the nearest integer, and $\Delta\mathbf{k}$ is a symbolic representation of the integer coordinate offsets needed to reach some close-range neighborhood of $[T(\mathbf{k})]$. The PV method puts emphasis on the weights of this sum. More precisely, the joint histogram can be expressed as

$$h[\phi, \gamma] = \sum_{\mathbf{k}_1 \in \mathbb{K}_1} \delta[f(\mathbf{k}_1) - \phi] \sum_{\mathbf{k}_2 \in \mathbb{K}_2} \delta[g(\mathbf{k}_2) - \gamma]\, \beta^1(T(\mathbf{k}_1) - \mathbf{k}_2), \tag{5}$$

where $\mathbb{K}_1 = \mathbb{K}_2 = \mathbb{D}$ is the domain of overlap of the two images, and where $\delta(x) = 1 - |\mathrm{sgn}(x)|$ is the Kronecker delta. In (5) we observe that only terms of the form $g(\mathbf{k})$ are computed, but that no term of the form $g(T(\mathbf{k}))$ appears. Consequently, for this approach to the computation of $h$ to be well-behaved, it is necessary that $\mathbb{D} \subseteq \mathbb{Z}^q$, with $q = 2$ for images, and $q = 3$ for volumes. In other words, all samples must necessarily be chosen at integer coordinates, a requirement which is still present for the related technique known as "Generalized Partial Volume," where $\beta^1$ in (5) is replaced by splines of higher degree [4]. Meanwhile, the discrete set $\mathbb{F}$ must be identical to the range of the quantized image $f$, and likewise for $\mathbb{G}$ and $g$. (Some freedom in the specification of $\mathbb{F}$ and $\mathbb{G}$ can be recovered by re-quantizing the images, whether linearly or not [3].) One component of the corresponding gradient can then be expressed as

$$\begin{aligned}\frac{\partial h}{\partial a} &= \sum_{\mathbf{k}_1 \in \mathbb{K}_1} \delta[f(\mathbf{k}_1) - \phi] \sum_{\mathbf{k}_2 \in \mathbb{K}_2} \delta[g(\mathbf{k}_2) - \gamma] \\ &\quad \cdot \left( \left. \frac{\mathrm{d}\beta^1(\mathbf{x})}{\mathrm{d}\mathbf{x}} \right|_{\mathbf{x} = T(\mathbf{k}_1) - \mathbf{k}_2} \right)^{\top} \frac{\partial}{\partial a} T(\mathbf{k}_1; \mathbf{a}).\end{aligned} \tag{6}$$

In practice, $h$ is never computed as the full double sum implied by (5) or (6). The preferred approach is to pick $\mathbf{k}_1 \in \mathbb{K}_1$ and to determine the small set $\mathbb{K} = \mathbb{K}_2 \cap \{\mathbf{k} : \beta^1(T(\mathbf{k}_1) - \mathbf{k}) \neq 0\}$. Then, it is sufficient to let $\mathbf{k}_2$ range over $\mathbb{K} \subseteq \mathbb{D}$ instead of over the whole set $\mathbb{K}_2 \subseteq \mathbb{D}$.

# 4 Halton Sampling

## 4.1 Low-Discrepancy Sequence

To perform random sampling, it is necessary to have a series of random coordinates at our disposal. Here, we show that some distributions are better than others for computing the mutual information. For example, let us consider the uniform distribution, which is the simplest of all and which is readily available in

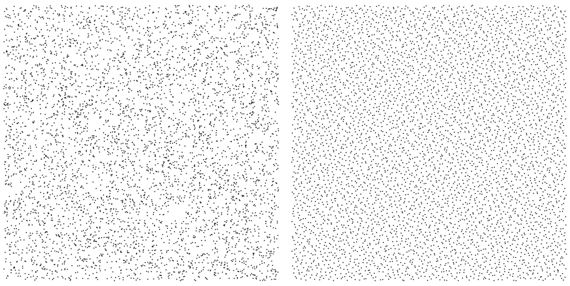

Figure 1: Random distribution of 5,000 dots. Left: uniform. Right: Halton sequence.

virtually every computer language. Many implementations are possible; a good discussion of some of them is available in [18]. In the left of Figure 1, we show 5,000 coordinates that are uniformly distributed within a square.

It is apparent from the left part of Figure 1 that the square is not filled out as uniformly as expected. Clustering effects are obvious; conversely, some spots are not covered. The quality of uniformity of a point set of size $N$ in $q$ dimensions is called "Discrepancy" [7, 44]. It is given by

$$\begin{aligned} D_N^*(\{\mathbf{x}_0,\ldots,\mathbf{x}_{N-1}\}) &= \sup_{0<v_i\leq 1, i\in\{1,\ldots,q\}} \left| \prod_{i=1}^{q} v_i \right. \\ &\qquad \left. - \frac{\operatorname{card}(\prod_{i=1}^{q} [0,v_i[ \cap \{\mathbf{x}_0,\ldots,\mathbf{x}_{N-1}\})}{N} \right|, \end{aligned}$$

where, without loss of generality, the coordinates $\mathbf{x}$ belong to the unit cube $[0,1[^q$. This expression measures by how many the number of points in a subvolume deviates from the number of points that would be expected under an isorepartition assumption. The worst deviation is of interest here.

A low-discrepancy sequence is such that its discrepancy decays asymptotically at least as fast as $\mathcal{O}((\log N)^q/N)$, where $q$ is the dimension. A Van der Corput sequence has this property for $q = 1$; it can be constructed easily by mapping a sequence of $N = (K - k)$ consecutive nonnegative integers $\mathbb{U} = \{k, k+1, \ldots, K-1\}$ into a sequence of rational numbers $\mathbb{V} = \{r_k, r_{k+1}, \ldots, r_{K-1}\}$, such that each member of the new sequence is the radical inverse of the corresponding member in the original sequence [7]. We write

$$r_n = B^{-L} \sum_{l=0}^{L-1} b_{L-1-l}\, B^l, \tag{7}$$

where $\{b_0, b_1, \ldots, b_{L-1}\}$ are the digits of the expansion of $n$ in base $B \in \mathbb{N}\backslash\{0\}$, with $L = 1$ for $n = 0$ and $L = 1 + \lfloor \log_B n \rfloor$ for $n > 0$. More precisely,

$$n = \sum_{l=0}^{L-1} b_l \, B^l, b_l \in \{0, 1, \ldots, B-1\}.$$

Obviously, the inspection of (7) reveals that $r_n \in [0, 1[$ for $n \in \mathbb{N}$. Moreover, it ensures that all samples $r_n$ are relevant because none can ever be taken twice—they all have a different value.

A Halton sequence [7, 14] is the multidimensional equivalent of Van der Corput sequences; it is obtained by mapping the original unidimensional integer sequence $\mathbb{U}$ into the multidimensional rational sequence $\mathbb{V}^q$, using a different base for each dimension. These bases must be relative primes; as the asymptotic regime is reached earlier for small primes, it is recommended to use the smallest primes available. We show in the right part of Figure 1 a series of 5,000 coordinates that correspond to a Halton sequence with $B_1 = 2$ and $B_2 = 3$. Visually, the distribution of the set of coordinates that are shown in the right part of Figure 1 is better spread out than that in the left part of the same figure, which confirms the theoretical predictions about the discrepancy of the respective sequences. Many other constructions of low-discrepancy sequences are known [9, 10, 15, 34], but they are more complicated to generate [19], and they play a significant role only when the number of dimensions is much larger than what we consider here.

## 4.2 Computational Cost

With respect to computational demands, the main disadvantage of the proposed method over previously published ones is its need for interpolation of the target image $f$, in addition to the traditional interpolation of the source image $g$. However, since the sequence of $(K - k)$ coordinates $\mathbf{x}_n \in \mathbb{V}^q$ is uniquely determined, the sequence of values $\{f(\mathbf{x}_n)\}$ can be precomputed. When $\mathbf{x}_n \notin \mathbb{D}$, replacement realizations are required; those can either be produced on demand, or precomputed, too, for example, by ensuring that $(K-k)$ exceeds the prescribed number $N_D$ of realizations by a sufficient margin.

The cost of high-quality interpolation, which we believe is necessary to obtain high-quality registration, is much higher than that of all other contributions. For example, the use of cubic-spline interpolation in 3-D requires the computation of $4 \cdot 3 = 12$ values of a basis function; those values are then used to build a weighted sum of $4^3 = 64$ image samples. By contrast, we estimate the cost of generating one Halton-based random coordinate $\mathbf{x}_n$ to be equivalent to the cost of computing a couple of values of a basis function. Similarly, the cost of computing the geometric transformation $T(\mathbf{x}_n)$ is negligible. Because $\mathbb{V}^q$ does not correspond to a raster array, we cannot benefit from acceleration techniques that take advantage of a structural interaction between a regular grid of samples and some types of global transformation (*e.g.*, rigid-body, affine, polynomial), but the resulting loss in efficiency is not relevant because most of the computation time is spent in interpolation anyway.

# 5 Experimental Validation

In Sections 5.1–5.3, we first illustrate the problems that typically arise with the use of *regular* sampling. To isolate the cause of each artifact as clearly as possible, we paint this illustration gradually, starting by a bland canvas (Section 5.1), and proceeding stroke by stroke to introduce the grid effect and the overlap problems in Sections 5.2 and 5.3, respectively. Then, we introduce irregular sampling in Section 5.4–5.7, and show that it removes every blemish of our illustrative example. From Section 5.8 on, we address the issue of objectively validating our approach.

## 5.1 Ideal Case

We start by performing a simple experiment to investigate the behavior of the mutual information under ideal conditions. We take the classical $(512 \times 512)$ Lena image in the role of both $f$ and $g$, and let the transformation $T(\mathbf{x}; 0, \Delta y)$ be a vertical translation of $\Delta y$ pixels, which amounts to sliding up and down the source image $g$ over a fixed target image $f$. Restricting this translation to the range $[-2.5, 2.5]$ pixels, we ensure that the domain of overlap $\mathbb{D}$ remains constant by setting to 1 each of the $(512 \times 512)$ elements of $G$ and each of the $(512 \times 512)$ elements of $F$, except for a 5 pixel-wide margin on the four sides[1] of $F$, which results in an effective size $(502 \times 502)$ for $\mathbb{D}$. Thus, the pristine mask $G(T(\mathbf{x}; 0, \Delta y))$ slides over the eroded mask $F$. We present the resulting plot of $I$ *vs.* vertical translation $\Delta y$ in the left part of Figure 2. The dotted curve corresponds to the Partial Volume (PV) approach, while the solid curve corresponds to the Spline-Parzen (SP) approach. Both assume that the joint histogram needed to compute (1) has been obtained by regular sampling of the images. Both satisfy $\mathrm{card}(\mathbb{F}) = \mathrm{card}(\mathbb{G}) = 200$, an arbitrary choice. We observe a well-defined maximum at $\Delta y = 0$, a translation for which $f = T\{g\}$. One could deduce that mutual information is always well-behaved near the optimal registration, but, as we shall see shortly, this is not true.

## 5.2 Grid Effect

The conditions that lead to the left part of Figure 2 were exceptionally favorable, in the sense that the trajectory in the $(\Delta x, \Delta y)$-space of parameters was reaching

---

[1] The width of this margin is arbitrary. The horizontal component of the margin will come useful for the experiments of Section 5.2.

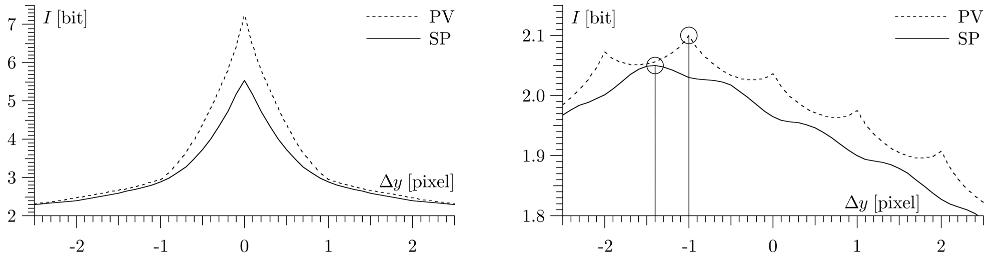

Figure 2: Mutual information $I$ *vs.* vertical translation $\Delta y$ of the Lena image for the partial volume (PV) and spline Parzen (SP) methods of computation of a joint histogram. Left: Perfect registration is obtained for $\Delta y = 0$, which coincides with the maximum of $I$. Right: The horizontal translation $\Delta x = 2$ is enforced. The grid effect is less pronounced for SP than for PV. Both: The area of the domain of overlap remains constant.

the global optimum. In practice, we are almost never so lucky and the trajectory often misses the global optimum. Thus, in order to perform an experiment that is more representative of a real registration task, where the global optimum is not reached directly, we now change slightly the conditions by enforcing and maintaining a constant horizontal offset of $\Delta x = 2$ pixels. In other words, we consider the translation $T(\mathbf{x}; \Delta x, \Delta y)$ with $\Delta x = 2$ instead of $\Delta x = 0$. We present the resulting plot of $I$ *vs.* vertical translation $\Delta y$ in the right part of Figure 2, where it is obvious that the slight change that we have introduced produces dramatic effects. The top curve (PV) now exhibits many local extrema, which exacerbates the difficulty of the optimization problem. Moreover, the pattern of maxima clearly coincides with the grid of samples over which $f$ is defined. The SP method suffers less from these artifacts which are called "Grid Effect"; nevertheless, they are not totally suppressed.

The experiments above have been conducted with a single modality, but the specific interest of the mutual information is to solve for registration problems where correlation methods fail—in other words, where $f$ and $g$ are very different. We now investigate just such a case and replace the Lena image by the the pair of brain volumes shown in Figure 3. When we perform a translation experiment along an axis perpendicular to the displayed slice, we observe a strong grid effect, as seen in the curves PV-co and SP-co of Figure 4. For this experiment, the domain of overlap remains constant—"co" stands for constant overlap, and "vo" for varying overlap. These curves have been obtained by making again the arbitrary choice $\mathrm{card}(\mathbb{F}) = \mathrm{card}(\mathbb{G}) = 200$. The PV method has many local maxima which are strongly biased toward integer values, while the SP method is biased toward half integers. We see that the bias is lesser for SP than for PV,

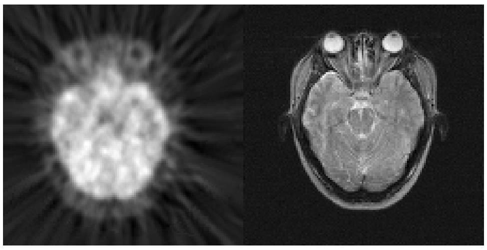

Figure 3: Left: Slice of a 3-D scan of a human brain acquired by Positron Emission Tomography (PET). Right: Slice of a 3-D scan of the same brain acquired by Magnetic Resonance Imaging (MRI-T2). Here, the alignment of these two slices is approximate only.

which reduces the risk of being trapped in a non-global optimum—over the range considered, PV-co exhibits five maxima out of five integer translations $\Delta z$, while SP-co exhibits only two objective maxima out of six half-integer translations $\Delta z$. But it would be even better if there would be no bias at all; we defer its complete removal until Section 5.4.

## 5.3 Overlap

By sliding a large $G$ over a small $F$, we were able to keep constant the domain of overlap between $f$ and $g$ in Sections 5.1 and 5.2. We could control the overlap because we knew in advance the maximal extent of the translation, but this knowledge would not have been available in the uncontrolled situation of a real registration task.

By setting now $F = G = 1$ over the support of the whole image, and by sliding $G$ under $F$, we obtain that the overlap is maximal when $T$ is the identity transformation, and that it decreases in step with the translation. This results in the curves PV-vo and SP-vo shown in Figure 4, where we observe a severe loss of regularity in the SP case—the overlap problem—in addition to the grid effect. For example, a discontinuity of the SP-vo curve is clearly visible for $\Delta z = 0.5$; optimization algorithms that rely on a local model of the similarity measure, such as [39], do not fare well under those conditions and require multiresolution to bypass this issue.

The overlap problem is especially pronounced in the present experiment because the translation direction is such that the volume of overlap varies by increments of a whole slice at a time. As the volume consists of few slices only, the ratio $\Delta\text{card}(\mathbb{D})/\text{card}(\mathbb{D})$ is substantial. Meanwhile, the PV curve remains continuous, due to the use of a soft-edge mask. Although this last improvement is applicable to many methods of computation of the joint histogram, we have not applied it to SP in this paper because the much larger support of

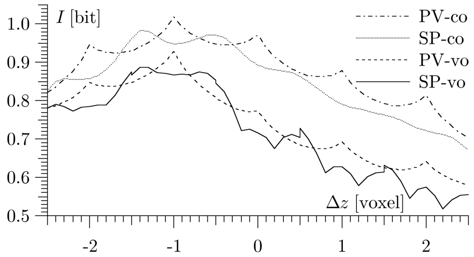

Figure 4: Mutual information $I$ *vs.* translation $\Delta z$ of the volumes shown in Figure 3, perpendicularly to the slices. The curves PV-co and SP-co have been built with a constant volume of overlap. A varying overlap characterizes the curves PV-vo and SP-vo. The volume of the domain of overlap varies gradually in the partial volume (PV-vo) case and varies stepwise in the spline Parzen (SP-vo) case—the transitions happen when $\Delta z$ is half-integer. The overlap problem manifests itself as discontinuities in the curve $I(\Delta z)$.

the basis function introduces additional complexity which we are not willing to deal with—while there are only eight neighbors to consider for PV, there are sixty-four for SP. In Section 4, we propose an alternative solution to the overlap problem that restores continuity, even under the difficult conditions of these experimental settings.

### 5.4 Stochastic Sampling

We now remove the assumption of regular sampling inherent in the SP approach. Irregular sampling does away with the grid effect and is extremely beneficial to the robustness and accuracy of registration; the cost is that interpolation is needed to know the value of the sample $f(\mathbf{x})$, in addition to that of $g(T(\mathbf{x}))$, since the coordinate $\mathbf{x}$ is not necessarily integer anymore.

In the context of the SP method, this results in Figure 5, where the conditions are exactly the same as those of the SP-vo curve in Figure 4, but for the fact that the coordinates $\mathbf{x}$ of the samples used to build the joint histogram $h$ are now realizations of a 3-D random variable. Those realizations of $\mathbf{x}$ that fall outside the domain of overlap of $F$ and $T\{G\}$ are rejected, and replacement realizations are produced until a prescribed number $N_D$ of realizations is reached; this procedure is responsible for the strong attenuation of the discontinuities that were present in the curve SP-vo of Figure 4. Obviously, the curves in Figure 5 do suffer much less from non-global maxima than the curves in Figure 4, and are easy to optimize, particularly near the expected optimum for the translation that we considered.

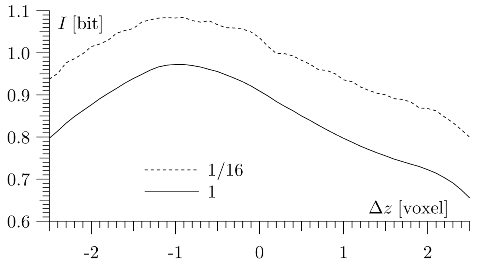

Figure 5: Mutual information $I$ *vs.* translation $\Delta z$ of the pair of volumes shown in Figure 3, perpendicularly to the slices. The volume of the domain of overlap varies stepwise—the transitions happen when $z$ is half-integer—but the number of samples remains constant. The curve is much smoother than those in Figure 4 because of random sampling. The solid curve has been obtained by using as many samples as there are voxels in the target volume; the dotted curve has been obtained by reducing this number of samples by a factor sixteen.

## 5.5 Number of Samples

Stochastic sampling offers the freedom to specify an arbitrary number of samples $N_D$. On one hand, reducing the number of samples can accelerate the computations. On the other hand, since we have introduced a random process, our similarity measure is not deterministic anymore; a reduction in $N_D$ leads to a greater variance $\mathbf{VAR}\{I\}$, and to potential aliasing.

We show in Figure 5 what happens if we let the nominal $N_D$ be reduced by a factor sixteen. We see that the change in variance is not the only effect; the amount of mutual information (*i.e.*, the measure of the quality of prediction) also increases when reducing $N_D$, because it is easier to predict few samples than many. This has for consequence that the use of a prestored set of random coordinates comes at a risk since comparing $I\{f, T_1\{g\}\}$ to $I\{f, T_2\{g\}\}$ may introduce a *systematic bias* when $T_1$ and $T_2$ are such that $N_{D_1} \neq N_{D_2}$, which is a situation that typically occurs for a change of overlap.

## 5.6 Variance Reduction

We have estimated the empirical standard deviation of $I$ for the plain curve in Figure 5 by computing 100 realizations of this curve. We have considered two cases. In the first one, we build the joint histogram $h$ by choosing random coordinates that follow a uniform distribution. In the second case the random coordinates are members of a Halton sequence in which a random choice of the first member $k$ of $\mathbb{U}$ ensures that a different sequence of consecutive integers $\mathbb{U} \subset \mathbb{N}$ is used for every new estimate of $h$.

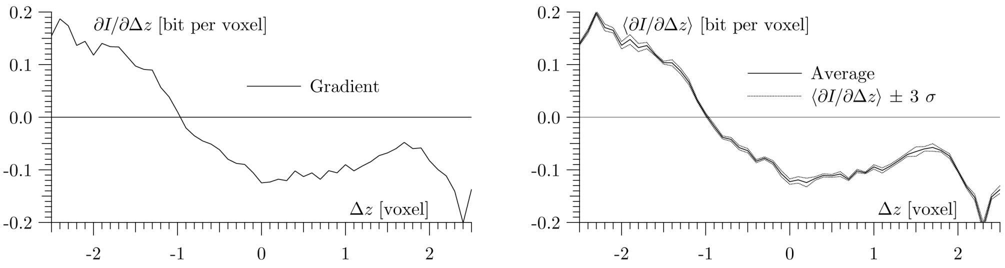

Figure 6: Left: Single realization of the gradient of the mutual information $\partial I/\partial \Delta z$ *vs.* translation $\Delta z$ of a pair of volumes for random sampling based on Halton sequences. Right: Mean gradient of the mutual information $\langle \partial I/\partial \Delta z \rangle$ *vs.* translation $\Delta z$ of a pair of volumes for random sampling based on Halton sequences. The dotted curves indicate ±3 standard deviations. The plain curve is the mean of 100 realizations.

When the random coordinates follow a uniform distribution, the largest standard deviation over all curves and all translations is $\max(\sigma) = 0.0024$, which is small when compared to the range of $I$ observed in Figure 5. Despite its already diminutive size, this standard deviation can be further reduced by carefully controlling the randomness of $\mathbb{D}$. For spaces that possess a low number of dimensions (images or volumes), the use of Halton sequences can substantially abate the variance of a quantity estimated by stochastic sampling [14]. In the present case, at least a five-fold reduction of the standard deviation has been observed. The largest standard deviation over 100 realizations then reduces to $\max(\sigma) < 0.00039$ (worst case). Meanwhile, the median of all standard deviations reduces more than tenfold, from 0.0019 to 0.00017.

Because of random sampling, we are facing the task of optimizing a nondeterministic similarity measure $I$. The best approach would be to fully acknowledge the existence of the random component of $I$, which typically would lead to designing an optimizer that must evaluate $I(\mathbf{a})$ several times over for any given parameter $\mathbf{a}$, so as to obtain a reliable estimate of the expected value $\mathbf{E}\{I(\mathbf{a})\}$. Unfortunately, this approach comes at an overly high computational cost. We propose instead to ensure that the variations due to the randomness of $I$ have a very small amplitude, which is precisely what the use of a Halton sequence offers. Practically, we can then safely disregard the random component of the similarity measure. The optimization procedure remains deterministic and gains in efficiency because it is sufficient to evaluate $I(\mathbf{a})$ once only.

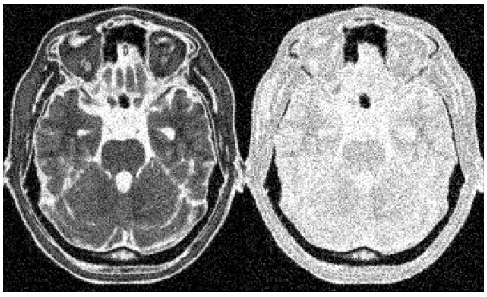

Figure 7: Left: One of the target image $f$ is the forty-fifth slice of the T2 modality of the simulated brain database with 10 dB added noise. Right: One of the source images $g$ is the forty-fifth slice of the PD modality of the simulated brain database with 10 dB added noise.

## 5.7 Gradient

It is often possible to boost the performance of an optimizer when information about the gradient is made available. In our case, we never estimate $\partial I/\partial \mathbf{a}$ by a finite-difference approach, but we use instead the analytical forms of this gradient developed in (4) and (6). This leads to a gradient that is the realization of a random variable (which exhibits a small variation), as opposed to the "gradient" of the realization of a random variable—which could be realized by a central-difference scheme but would be erratic and meaningless. We have used cubic splines throughout this paper to build the data model, except in the cases that involve PV, for which no explicit interpolation takes place.

The component of the analytical gradient of the bottom curve in Figure 5, in the direction of the displacement, is given in the left part of Figure 6. We see that this curve has a single zero-crossing that neatly coincides with the maximum in Figure 5. We observe that this maximum is near $\Delta z = -0.95$, but it is unrealistic to assume that this translation corresponds to the best overall alignment because other parameters such as rotations and translations along the remaining axes should be jointly optimized to reach the global optimum. In the present case, only translation along $z$ is considered, whereas the other parameters do not take their optimal value. This situation is akin to that encountered in the right part of Figure 2. Repeating the estimation of this curve for 100 different realizations of the Halton sequence for each translation $\Delta z$, we obtain the mean curve and its standard deviation that we show in the right part of Figure 6. We observe that the largest standard deviation over 100 realizations is less than $\max(\sigma) = 0.0031$ (worst case).

### 5.8 Objective Validation

We perform now the following experiment to validate our method objectively: We start with two images $f$ and $g$ that are already in perfect registration, albeit with a nonlinear and noninvertible mapping between the intensities; we then perturb this alignment by applying by cubic-spline interpolation a known transformation $T_i$ of parameters $\mathbf{a}_i$ to the second image, which becomes the perturbed source image $g_i' = T_i\{g\}$. Adding noise results in $f_i$ and $g_i$. The principle of the registration task is then to register $g_i$ to $f_i$, using the identity as initial guess for the iterative optimization algorithm. When the estimated registration parameters $\hat{\mathbf{a}}_i$ are close enough to the true $\mathbf{a}_i$, the registration is said to be successful; else, it is said to have failed. The set of parameters $\hat{\mathbf{a}}_i$ for which the registration is almost always successful is called the capture range of the registration algorithm.

We compare the capture range of the method proposed in this paper to that of the algorithm presented in [39]. The one major difference between the two approaches is the set $\mathbb{D}$ of coordinates from which the joint histogram is estimated. In the present paper, it is given by a Halton sequence, while in [39] the coordinates are deterministic and sit on the nodes of a regular Cartesian lattice.

The Simulated Brain Database[2] presents us with convenient data consisting in three modalities (PD, T1, T2) that, being simulated, are in perfect coregistration [20]. For the present experiment, we consider all six pairings of modalities, noting that the attempt to register data from modality A to data from modality B may not be experimentally equivalent to registering B to A. To achieve better statistical significance, instead of registering a single 3-D volume we have multiplied further the number of cases by performing experiments over several 2-D slices independently; more precisely, from a $(181 \times 217 \times 181)$ volume we have retained the five $(181 \times 217)$ slices indexed $\{35, 40, 45, 50, 55\}$. Then, we have conducted 200 experiments for each pair of modalities and for each slice, for a total of 6,000 experiments per method. To make the task more challenging, for each experiment $i$ we have added to both $f$ and $g_i'$ a new independent realization of noise $n_{f,i}$ and $n_{g,i}$, so that $g_i = g_i' + n_{g,i} = T_i\{g\} + n_{g,i}$ and $f_i = f + n_{f,i}$. This added noise, which follows a normal distribution and which has been added after the application of the transformation $T_i$ so as to keep it uncorrelated, is such that $\mathbf{VAR}\{n_f\} = 0.1\,\mathbf{VAR}\{f\}$ and $\mathbf{VAR}\{n_g\} = 0.1\,\mathbf{VAR}\{g\}$. This corresponds to 10 dB signal-to-noise ratio. We illustrate a pair of noisy slices in Figure 7.

To simplify the interpretation of the results of our experiments, we have chosen to work with rigid-body transformations. The angle of rotation $\theta_i$ is uniformly distributed in $[-10°, 10°]$, while the translation $\{\Delta x_i, \Delta y_i\}$ is uniformly distributed in $[-10, 10]$ pixel. The estimated registration parameters $\hat{\mathbf{a}}_i$ can then be represented in homogenous coordinates as the matrix $\hat{\mathbf{G}}_i$, while the true transformation matrix is $\mathbf{G}_i$. We report the success of the registration as a quantity that we call the mean-square warping index $\varpi^2 = \mathrm{trace}((\mathbf{G} - \hat{\mathbf{G}})^\top (\mathbf{G} - \hat{\mathbf{G}}))$, the initial mean-square warping index (before registration) being given by $\varpi_0^2 = \mathrm{trace}((\mathbf{G} - \mathbf{I})^\top (\mathbf{G} - \mathbf{I}))$. The precise meaning of $\varpi^2$ is detailed in the Appendix.

---

[2] http://www.bic.mni.mcgill.ca/brainweb/

We have explored six registration methods. Among them, the traditional partial-volume method (PV), the single-resolution spline-based Parzen-window (SRSP), and the multiresolution spline-based Parzen-window (MRSP) provide a baseline to which our proposed approach (HSP) and its variations (HSPO and USP) can be compared.

With critical sampling (as many samples as there are pixels in the target image) and with $\mathrm{card}(\mathbb{F}) = \mathrm{card}(\mathbb{G}) = 100$, the method of this paper (Halton-based Spline Parzen, HSP) results in subpixel registration all of the time. When pooled over all 6,000 experiments, the warping index of our proposed method is $\frac{1}{6000} \sum_{i=1}^{6000} \varpi_i^2 = 0.0045$, with a maximum $\max_{i=1}^{6000}(\varpi_i^2) = 0.17$. This result has been obtained without multiresolution, and no outlier was observed.

Our previous algorithm (SP) required multiresolution to achieve robustness (Multi-Resolution SP, or MRSP). Under the same conditions as above, but with regular sampling and a five-level pyramid, we observe that the registration reaches a subpixel accuracy 85% of the time. This is somewhat better than the traditional PV approach, which succeeds only 66% of the time. These imperfect results testify that the current registration problem is a difficult one due to the presence of heavy noise. We had explained in [39] that SP was designed to work hand-in-hand with a multiresolution pyramid. If the multiresolution stage is skipped and registration is attempted on full-scale data directly (SRSP), then we observe a host of outliers. Without multiresolution, our previous method often stalls right at the start of the optimization. By contrast, the method of this paper exhibits excellent robustness, even in the absence of a multiresolution pyramid.

We provide detailed results of these experiments in Tables 1 and 2. In the first table, we report the result of every experiment. There, it is interesting to note that some registration attempts are able to produce a final error that is much larger than the worst initial error $\max(\varpi_0) = 196.8$. This can happen because nothing prevents the solution to drift to an absurd position. When this mishap occurs, multiresolution ceases to be an ally because of the magnification of those erratic drifts at low resolution. In particular, the consequence of a failure at the coarse level of MRSP is sometimes spectacular, because the drift of the solution is amplified when the erroneous result is propagated to finer levels. This explains why, in the MRSP case, the mean-square warping index is much larger than $\max(\varpi_0)$. Meanwhile, SRSP stalls right at the beginning of the optimization, which results in its final mean-square warping index being

Table 1: Mean-square warping index $\varpi^2$ for all runs, including failures. HSP: Halton-Based Spline Parzen. HSPO: HSP with Overlap dependence. USP: Uniform sampling Spline Parzen. MRSP: Multi-Resolution Spline Parzen. PV: Partial Volume. SRSP: Single Resolution Spline Parzen.

| | Sampling | Failures | Mean ± Std. Dev. | Max |
|---|---|---|---|---|
| $\varpi_0^2$ | None | 5958 | 66.8 ± 42.6 | 196.8 |
| HSP | Halton | 0 | 0.0045 ± 0.0066 | 0.17 |
| HSPO | Halton | 200 | 194.0 ± 1452.7 | 18971.4 |
| USP | Uniform | 1 | 0.0064 ± 0.0561 | 4.3 |
| MRSP | Regular | 912 | 837.0 ± 4629.7 | 39663.1 |
| PV | Regular | 2070 | 17.2 ± 40.5 | 196.8 |
| SRSP | Regular | 5921 | 65.4 ± 43.3 | 196.8 |

barely different from the initial one. As the contribution of these extreme cases dominate when computing mean and standard deviation, the results of Table 1 is not a fair representation of the intrinsic performance of the methods. Therefore, we report in Table 2 the result of those experiments that produced a subpixel final error. We observe that HSP performs best, both in terms of accuracy and robustness.

We have provided two suboptimal variations of our methods (HSPO and USP) to gain more insights into the approach of this paper. In the first case (HSPO) we wanted to highlight the magnitude of the overlap problem. To that effect, we have built a registration method that is exactly identical to HSP in every respect, but for the fact that no additional realization of coordinates has been produced whenever $\mathbf{x}_n \notin \mathbb{D}$. Thus, the number of samples effectively depends on the area of overlap for HSPO. Under such conditions, the robustness suffers since 200 non-subpixel occurrences are observed whereas there were none before. We attribute this loss in robustness to the fact that the optimizer builds a pointwise Taylor-like model of the similarity measure [39], which is usually a desirable property (being pointwise, this model enjoys a better accuracy than constructions based on finite differences), but which is bound to fail near the discontinuities resulting from the overlap dependence since the domain of convergence of the pointwise model does not extend beyond them. In addition, we observe that the final accuracy is reduced too, even in the subpixel cases.

The second additional method (USP) helps us to investigate the effect of the sampling pattern. USP is identical to HSP in every respect, including the production of additional realizations of coordinates when needed, but differs by the fact that samples drawn from a uniformly distributed probability density function have been used instead of using Halton sampling. The effect on robust-

Table 2: Subpixel Mean-Square warping index $\varpi^2$ for successful runs only. HSP: Halton-Based Spline Parzen. HSPO: HSP with Overlap dependence. USP: Uniform sampling Spline Parzen. MRSP: Multi-Resolution Spline Parzen. PV: Partial Volume. SRSP: Single Resolution Spline Parzen.

| | Sampling | Successes | Initial | Final |
|---:|---:|---:|---:|---:|
| HSP | Halton | 6000 | 66.8 ± 42.6 | 0.0045 ± 0.0066 |
| HSPO | Halton | 5800 | 64.0 ± 40.4 | 0.0050 ± 0.0102 |
| USP | Uniform | 5999 | 66.8 ± 42.6 | 0.0057 ± 0.0083 |
| MRSP | Regular | 5088 | 61.3 ± 39.8 | 0.0069 ± 0.0364 |
| PV | Regular | 3930 | 61.0 ± 37.6 | 0.0947 ± 0.2034 |
| SRSP | Regular | 79 | 1.5 ± 1.9 | 0.4416 ± 0.2916 |

ness is negligible, but not the effect on accuracy, which confirms the benefit of using a low-discrepancy sampling pattern, as we saw already in Section 5.6.

### 5.9 Validation on Clinical Data

We have applied our algorithm to the registration of volumes acquired by CT (Computed Tomography) or PET (Positron Emission Tomography) with respect to three different MRI (Magnetic Resonance Imaging) modalities: PD (Proton Density), T1 (T1 Relaxation Time), and T2 (T2 Relaxation Time). The goal was to align the CT or the PET volumes with the MRI ones. The MRI volumes were available in two versions: uncorrected (PD, T1, T2) and corrected for scanner-dependent geometric distortion (PDr, T1r, T2r). There were up to seven patients in most cases.

We compare the results of our intermodal brain image registration algorithm to those of several other approaches. The comparison is based on a methodology proposed by West *et al.*, who let researchers access a standard set of volumes to be registered [41, 42, 43]. They also act as a repository for the ideal registration transformations (gold standard) acquired by a prospective method using fiducial markers. These markers are erased before the volumes are disclosed to the investigators, who then face a retrospective blind registration task. (We were blinded in exactly the same way as the researchers listed in [42], and still are.) After registration, the investigators report back a set of transformation parameters that are compared to the gold standard. This results in a geometric error measured in mm, and allows for a simple ranking of the competing algorithms—from an accuracy point of view. The errors are measured at 10 specific locations of clinical interest. Those locations that happen to lie outside of the field of view for some of the volumes are discarded from the evaluation.

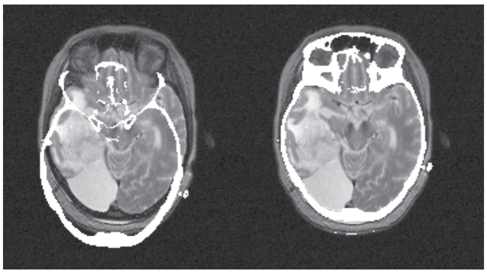

Figure 8: Registration in 3-D of a CT to a rectified T2. Left: initial correspondence. Right: final registration.

We show in Figure 8 an example of registration of a CT to a rectified T2. We have shown the bones from the CT as bright saturated features that overlay the T2. Before registration, the initial fit is clearly off, both within slice and in the direction perpendicular to the slice. After registration, the CT has been properly aligned to the T2 and no multiresolution optimization was required to achieve this apparently good fit. For this specific patient and for the CT-T2r case, the median error related to the gold standard is 0.85 mm.

We give in Tables 3 and 4 the summary of all results obtained by our algorithm. We present the mean, the median, and the maximum error for each pair of modalities. The number of patients is $P$. We have also provided the standard deviation $\sigma$ of our errors—outliers included. For comparison purposes, the numbers in parentheses correspond to the results of our previous algorithm [39] which was computing the similarity measure with a spline Parzen method, too, but with regular sampling.

These results are representative of the accuracy of our method only if the gold-standard provides perfect registration; unfortunately, this is impossible to obtain in a physical world. The authors of [42] have conducted an error and sensitivity analysis of their data. Based on simulations, they reach the conclusion that the root-mean-square error of the gold standard itself is about 1.65 and 0.39 mm for PET-to-MR and for CT-to-MR cases, respectively. In addition, they claim that the standard deviation of registration methods that achieve an accuracy similar to that of the gold standard is expected to be around 2.33 and 0.55 mm for PET-to-MR and for CT-to-MR cases, respectively.

In the case of PET-to-MR registration we see that our score for the median error nears the assumed accuracy of the gold standard (1.65 mm), particularly for cases involving the T2 or T2r modality. Except in the PET-T1 case, all our scores are within one millimeter of this limit. At the time of writing this paper, we know of no other method that could reach a better agreement with the gold standard in the PET-T2 case over the whole data set; our results in the PET-PD case are also the best obtained so far. In the PET-T2r case we

Table 3: Residual difference in mm between a fiducial-marker prospective registration technique and the proposed retrospective algorithm (HSP) for PET-MRI modalities. Parenthesized results: multiresolution regular sampling [27].

|  | PET-T1 | PET-PD | PET-T2 | PET-T1r | PET-PDr | PET-T2r |
|---|---|---|---|---|---|---|
| $P$ | 7 | 7 | 7 | 4 | 5 | 5 |
| Mean | 4.18 | 3.23 | 2.31 | 2.48 | 2.80 | 2.17 |
|  | (4.52) | (3.37) | (2.83) | (2.16) | (2.93) | (2.59) |
| Median | 3.07 | 2.66 | 1.85 | 2.54 | 2.47 | 1.73 |
|  | (3.10) | (2.68) | (2.61) | (1.89) | (2.05) | (1.96) |
| Max | 16.40 | 9.38 | 6.13 | 4.65 | 10.25 | 6.23 |
|  | (14.34) | (9.65) | (5.95) | (4.85) | (9.60) | (8.44) |
| $\sigma$ | 3.81 | 2.24 | 1.31 | 1.04 | 2.14 | 1.38 |
|  | (3.85) | (2.17) | (1.40) | (1.07) | (2.12) | (1.97) |

are second only to Rohlfing [32], who reached the median error 1.57 mm. The same investigators reached 1.56 mm in the PET-T1r case; both values are below the expected accuracy of the gold standard. The other best scores are from Woods [45] (PET-T1, 2.35 mm) and from Nikou [27] (1.93 mm, PET-PDr). These comparisons include only the investigators who processed the data for all $P$ patients; this still represents more than fifty complete submissions for each of the twelve pairs of modalities [11].

In the case of CT-to-MR registration no automatic retrospective registration method could as yet reach the assumed accuracy of the gold standard (0.39 mm). In examining first the results for uncorrected volumes, we observe that we nearly match the CT-T2 score that we obtained in [39], which is the best score reported so far by all investigators; despite this encouraging result, our accuracy is still about three times worse than the inferred accuracy of the gold standard. Qin [29] obtains the best scores on the CT-T1 uncorrected volumes (0.85 mm), and Tanács [36] reaches 1.51 mm in the CT-PD case. Not surprisingly, the best results available at the time of writing this paper are obtained for the three rectified volumes (T1r, PDr, T2r). By working with those volumes, we clearly obtain submillimeter scores. A comparison of our results to those of other investigators reveals that we are within 0.15 mm of the best contenders. Chen [4] obtained 0.62 mm in the CT-T1r case, while Ren [30] obtained 0.68 and 0.77 mm in the CT-PDr and CT-T2r cases, respectively.

To summarize, the accuracy of the proposed method as reported in Tables 3 and 4 compares favorably with similar published results obtained by other groups on the same data, especially for CT-T2, PET-PD, and PET-T2. However, when comparing accuracy to variance, we observe that the latter is not negligible with

Table 4: Residual difference in mm between a fiducial-marker prospective registration technique and the proposed retrospective algorithm (HSP) for CT-MRI modalities. Parenthesized results: regular sampling [27].

| | CT-T1 | CT-PD | CT-T2 | CT-T1r | CT-PDr | CT-T2r |
|---|---|---|---|---|---|---|
| $P$ | 7 | 7 | 7 | 6 | 7 | 7 |
| Mean | 1.36 | 2.04 | 1.50 | 0.90 | 0.82 | 1.09 |
| | (1.56) | (1.90) | (1.38) | (1.05) | (0.99) | (1.07) |
| Median | 1.27 | 1.93 | 1.14 | 0.77 | 0.73 | 0.87 |
| | (1.49) | (1.74) | (1.11) | (0.90) | (1.01) | (0.91) |
| Max | 2.65 | 4.09 | 3.45 | 1.97 | 1.85 | 2.95 |
| | (2.87) | (4.18) | (4.17) | (3.11) | (1.64) | (2.78) |
| $\sigma$ | 0.50 | 0.68 | 0.77 | 0.48 | 0.38 | 0.71 |
| | (0.50) | (0.85) | (0.80) | (0.59) | (0.37) | (0.56) |

respect to the former. Hence, the accuracy reported by most other investigators who used the mutual information as similarity measure cannot be distinguished from ours when statistical significance is taken into consideration. It is, however, a fact that the best all-purpose methods are all based on the mutual information; we have shown in this paper how to avoid two problems that were previously associated with this similarity measure.

## 6 Conclusion

The geometric alignment of a source to a target volume requires an objective similarity measure to compute the quality of fit. Many experiments, including those presented in this paper, point out that the similarity measure called "Mutual Information" is an excellent candidate that is applicable to many types of multimodal registration. However, its traditional computation involves a regular sampling grid, which leads to artifacts that are called "Grid Effect." These artifacts are detrimental to robustness and impede subvoxel accuracy.

Known solutions involve irregular sampling based on uniformly distributed random coordinates around a voxel. Instead of using a uniform distribution, we have proposed in this paper to achieve irregular sampling by the way of a Halton sequence. Because a Halton sequence enjoys a property called low-discrepancy, it is better suited than previous methods. An added benefit is the possibility to prescribe the number of samples; by enforcing that this number remains constant irrespective of the size of the domain of overlap, we also avoid the "Overlap Problem" which is characterized by discontinuities in the similarity measure.

Because of the random component inherent in our method, the similarity measure is non-deterministic. Nevertheless, the resulting variance of the mutual information is negligible thanks to the use of Halton sequences. Therefore, the use of a deterministic optimizer is more legitimate than in previously published approaches. Experimental evidence on the Vanderbilt retrospective registration task shows that the proposed approach performs well, at least within the limits on accuracy assessment imposed by the gold standard. We consistently reach a median accuracy that compares favorably with that of the best results obtained by more than fifty other groups of investigators. In three cases, we obtain the best results at the time of writing this paper (PET-PD 2.66 mm, PET-T2 1.85 mm, CT-T2 1.14 mm).

## ACKNOWLEDGMENT

This work was supported by the Center for Biomedical Imaging of the Geneva-Lausanne Universities and the EPFL, as well as by the Hasler, Leenaards, and Louis-Jeantet foundations. The images and the gold-standard transformations of Section 5 were provided as part of the project, "Evaluation of Retrospective Image Registration," National Institutes of Health, 1 R01 NS33926-02, Principal Investigator, J.M. Fitzpatrick, Vanderbilt University, Nashville TN, USA.

## Appendix A: Warping Index

In $q$ dimensions, the squared registration error at a coordinate $\mathbf{x}$ is given by the quantity $\|\mathbf{G}\,\mathbf{x}-\hat{\mathbf{G}}\,\mathbf{x}\|^2$. We collect all errors by integrating them over a spherical (or circular) ball $\mathbb{B}$ centered on the origin of the coordinate system in which $\mathbf{G}$ and $\hat{\mathbf{G}}$ are expressed. Because we want to build a global error measure that depends on $\mathbf{G}$ and $\hat{\mathbf{G}}$ only, but not on the radius $R$ of the ball $\mathbb{B}$, we normalize the resulting quantity by the integral of the squared distance to the origin. This defines the warping index $\varpi^2$ as

$$\begin{aligned}\varpi^2 &= \frac{q\int_{\mathbf{x}\in\mathbb{B}}\left\|\mathbf{G}\,\mathbf{x}-\hat{\mathbf{G}}\,\mathbf{x}\right\|^2\,\mathrm{d}^q x}{\int_{\mathbf{x}\in\mathbb{B}}\|\mathbf{x}\|^2\,\mathrm{d}^q x}\\ &= \mathrm{trace}(\left(\mathbf{G}-\hat{\mathbf{G}}\right)^\top\left(\mathbf{G}-\hat{\mathbf{G}}\right)).\end{aligned}$$

We now show that $\varpi^2$ does not depend on the radius $R$ of the $q$-dimensional hyperball $\mathbb{B}$. We first compute the matrix $\int_{\mathbf{x}\in\mathbb{B}}\mathbf{x}\,\mathbf{x}^\top\,\mathrm{d}^q x=\rho_q\,\mathbf{I}_q$, where $\mathbf{I}_q$ is the identity matrix in $q$ dimensions, and where $\rho_q=\frac{2\sqrt{\pi^q}\,R^{q+2}}{q\,(q+2)\,\Gamma(q/2)}$. Then, the denominator of $\varpi^2$ is given by

$$\int_{\mathbf{x}\in\mathbb{B}}\|\mathbf{x}\|^2\,\mathrm{d}^q x=\int_{\mathbf{x}\in\mathbb{B}}\mathbf{x}^\top\,\mathbf{x}\,\mathrm{d}^q x=\int_{\mathbf{x}\in\mathbb{B}}\mathrm{trace}(\mathbf{x}\,\mathbf{x}^\top)\,\mathrm{d}^q x=q\,\rho_q.$$

Meanwhile, the numerator of $\varpi^2$ is

$$\begin{aligned}
&q \int_{\mathbf{x}\in\mathbb{B}} \left\| \mathbf{G}\,\mathbf{x} - \hat{\mathbf{G}}\,\mathbf{x} \right\|^2 \mathrm{d}^q x \\
&\quad = q \int_{\mathbf{x}\in\mathbb{B}} \mathbf{x}^\top \left(\mathbf{G} - \hat{\mathbf{G}}\right)^\top \left(\mathbf{G} - \hat{\mathbf{G}}\right) \mathbf{x}\,\mathrm{d}^q x \\
&\quad = q \int_{\mathbf{x}\in\mathbb{B}} \mathrm{trace}(\left(\mathbf{G} - \hat{\mathbf{G}}\right)^\top \left(\mathbf{G} - \hat{\mathbf{G}}\right) \mathbf{x}\,\mathbf{x}^\top)\,\mathrm{d}^q x \\
&\quad = q\,\mathrm{trace}(\left(\mathbf{G} - \hat{\mathbf{G}}\right)^\top \left(\mathbf{G} - \hat{\mathbf{G}}\right) \int_{\mathbf{x}\in\mathbb{B}} \mathbf{x}\,\mathbf{x}^\top\,\mathrm{d}^q x) \\
&\quad = q\,\mathrm{trace}(\left(\mathbf{G} - \hat{\mathbf{G}}\right)^\top \left(\mathbf{G} - \hat{\mathbf{G}}\right) \rho_q\,\mathbf{I}_q) \\
&\quad = q\,\rho_q\,\mathrm{trace}(\left(\mathbf{G} - \hat{\mathbf{G}}\right)^\top \left(\mathbf{G} - \hat{\mathbf{G}}\right)).
\end{aligned}$$

The conclusion follows.

## References

[1] K.S. Arun, T.S. Huang, and S.D. Blostein, Least-Squares Fitting of Two 3-D Point Sets, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **9**(5), 698–700, September 1987.

[2] L.G. Brown, A Survey of Image Registration Techniques, *ACM Computing Surveys*, **24**(4), 325–376, December 1992.

[3] J. Camp and R. Robb, A Novel Binning Method for Improved Accuracy and Speed of Volume Image Coregistration Using Normalized Mutual Information, in *Proceedings of the SPIE International Conference on Image Processing: Medical Imaging*, K.M. Hanson, Ed., San Diego CA, USA, February 22-25, 1999, SPIE, **3661**, 24–31.

[4] H.-m. Chen and P.K. Varshney, Registration of Multimodal Brain Images: Some Experimental Results, in *Proceedings of the SPIE International Conference on Sensor Fusion*, B.V. Dasarathy, Ed., Orlando FL, USA, April 3-5, 2002, SPIE, **4731**, 122–133.

[5] H.-m. Chen and P.K. Varshney, Mutual Information-Based CT-MR Brain Image Registration Using Generalized Partial Volume Joint Histogram Estimation, *IEEE Transactions on Medical Imaging*, **22**(9), 1111–1119, 2003.

[6] A.A. Cole-Rhodes, K.L. Johnson, J. LeMoigne, and I. Zavorin, Multiresolution Registration of Remote Sensing Imagery by Optimization of Mutual Information Using a Stochastic Gradient, *IEEE Transactions on Image Processing*, **12**(12), 1495–1511, 2003.

[7] M. Drmota and R.F. Tichy, Eds., *Sequences, Discrepancies, and Applications*, *Lecture Notes in Mathematics*, **1651**, Springer-Verlag, Berlin, 1997.

[8] P.A. van den Elsen, E.-J.D. Pol, and M.A. Viergever, Medical Image Matching—A Review with Classification, *IEEE Engineering and Medicine in Biology*, **12**(1), 26–39, 1993.

[9] H. Faure, Discrépance de suites associées à un système de numération (en dimension $s$), *Acta Arithmetica*, **XLI**, 337–351, 1982.

[10] H. Faure, Suggestions for Quasi-Monte-Carlo Users, in *Proceedings of the International Conference on Finite Fields, Coding Theory and Advances in Communications and Computing*, G.L. Mullen and P.J.-S. Shiue, Eds., Las Vegas NV, USA, August 7-10, 1991 (1993), *Lecture Notes in Pure and Applied Mathematics*, **141**, 269–278, M. Dekker.

[11] J.M. Fitzpatrick, http://www.vuse.vanderbilt.edu/ image/registration/, National Institute of Biomedical Imaging and Bioengineering, Project Number 8R01EB002124-03.

[12] J. Flusser and T. Suk, A Moment-Based Approach to Registration of Images with Affine Geometric Distortion, *IEEE Transactions on Geoscience and Remote Sensing*, **32**(2), 382–387, 1994.

[13] K. Fukunaga, Ed., *Introduction to Statistical Pattern Recognition*, Electrical Science, A Series of Monographs and Texts. Academic Press, 1972.

[14] J.H. Halton, On the Efficiency of Certain Quasi-Random Sequences of Points in Evaluating Multi-Dimensional Integrals, *Numerical Mathematics*, **2**, 84–90, 1960.

[15] J.M. Hammersley, Monte-Carlo Methods for Solving Multivariate Problems, *Annals of the New-York Academy of Sciences*, **86**, 844–874, 1960.

[16] D.L.G. Hill, C. Studholme, and D.J. Hawkes, Voxel Similarity Measures for Automated Image Registration, in *Proceedings of the SPIE International Symposium on Visualization in Biomedical Computing*, R.A. Robb, Ed., Rochester MN, USA, October 4-7, 1994, SPIE, **2359**, 205–216.

[17] B. Kim, J.L. Boes, K.A. Frey, and C.R. Meyer, Mutual Information for Automated Unwarping of Rat Brain Autoradiographs, *Neuroimage*, **5**(1), 31–40, 1997.

[18] D.E. Knuth, Ed., *Seminumerical Algorithms*, vol. 2 of *The Art of Computer Programming*, Addison-Wesley, third edition, 1998.

[19] L. Kocis and W.J. Whiten, Computational Investigations of Low-Discrepancy Sequences, *ACM Transactions on Mathematical Software*, **23**(2), 266–294, 1997.

[20] R.K.-S. Kwan, A.C. Evans, and G.B. Pike, MRI Simulation-Based Evaluation of Image-Processing and Classification Methods, *IEEE Transactions on Medical Imaging*, **18**(11), 1085–1097, 1999.

[21] B. Likar and F. Pernuš, A Hierarchical Approach to Elastic Registration Based on Mutual Information, *Image and Vision Computing*, **19**(1-2), 33–44, 2001.

[22] B.D. Lucas and T. Kanade, An Iterative Image Registration Technique with an Application to Stereo Vision, in *Proceedings of the Seventh International Joint Conference on Artificial Intelligence (IJCAI'81)*, A. Drinan, Ed., Vancouver BC, Canada, August 24-28, 1981, 674–679, Kaufmann, W.

[23] F. Maes, *Segmentation and Registration of Multimodal Medical Images: From Implementation and Validation to a Useful Tool in Clinical Practice*, Ph.D. thesis, Faculty of Applied Sciences, Katholieke Universiteit Leuven, Leuven, Belgium, May 27, 1998, Thesis no. U.D.C. 681.3*I4 (1998).

[24] J.B.A. Maintz and M.A. Viergever, A Survey of Medical Image Registration, *Medical Image Analysis*, **2**(1), 1–36, 1998.

[25] N.J.I. Mars and G.W. Arragon, Time Delay Estimation in Non-Linear Systems Using Average Amount of Mutual Information Analysis, *Signal Processing*, **4**(2-3), 139–153, 1982.

[26] C.R. Maurer Jr. and J.M. Fitzpatrick, A Review of Medical Image Registration, in *Interactive Image-Guided Neurosurgery*, R.J. Maciunas, Ed., 17–44. American Associate of Neurological Surgeons, Park Ridge IL, USA, 1993.

[27] C. Nikou, F. Heitz, J.-P. Armspach, I.-J. Namer, and D. Grucker, Registration of MR/MR and MR/SPECT Brain Images by Fast Stochastic Optimization of Robust Voxel Similarity Measures, *NeuroImage*, **8**(1), 30–34, 1998.

[28] C. Nikou, F. Heitz, and J.-P. Armspach, Robust Voxel Similarity Metrics for the Registration of Dissimilar Single and Multimodal Images, *Pattern Recognition*, **32**(8), 1351–1368, 1999.

[29] B.-J. Qin and T.-G. Zhuang, Similarity Measures in Voxel Intensity Based 3D Multi-Modal Medical Image Registration, *Journal of Shanghai-Jiaotong University*, **36**(7), 942–945, 2002.

[30] H.-P. Ren, H. Yang, S.-Z. Chen, and W.-K. Wu, Study of Three-Dimensional PET and MR Image Registration Based on Higher-Order Mutual Information, *Nuclear Science and Techniques*, **13**(2), 65–71, 2002.

[31] C.E. Rodríguez-Carranza and M.H. Loew, A Weighted and Deterministic Entropy Measure for Image Registration Using Mutual Information, in *Proceedings of the SPIE International Conference on Image Processing: Medical Imaging*, K.M. Hanson, Ed., San Diego CA, USA, February 23-27, 1998, SPIE, **3338**, 155–166.

[32] T. Rohlfing, J.B. West, J. Beier, T. Liebig, C.A. Taschner, and U.W. Thomale, Registration of Functional and Anatomical MRI: Accuracy Assessment and Application in Navigated Neurosurgery, *Computer-Aided Surgery*, **5**(6), 414–425, 2000.

[33] C.Ó. Sánchez Sorzano, P. Thévenaz, and M. Unser, Elastic Registration of Biological Images Using Vector-Spline Regularization, *IEEE Transactions on Biomedical Engineering*, **52**(4), 652–663, 2005.

[34] I.M. Sobol, On the Distribution of Points in a Cube and the Approximate Evaluation of Integrals, *USSR Computational Mathematics and Mathematical Physics*, **7**(4), 86–112, 1967.

[35] H.S. Stone, M.T. Orchard, E.-C. Chang, and S.A. Martucci, A Fast Direct Fourier-Based Algorithm for Subpixel Registration of Images, *IEEE Transactions on Geoscience and Remote Sensing*, **39**(10), 2235–2243, 2001.

[36] A. Tanács, K. Palágyi, and A. Kuba, A Fully Automatic Medical Image Registration Algorithm Based on Mutual Information, in *Proceedings of the Third Conference of PhD Students in Computer Science*, Szeged, Hungary, July 1-4, 2002.

[37] P. Thévenaz, U.E. Ruttimann, and M. Unser, A Pyramid Approach to Subpixel Registration Based on Intensity, *IEEE Transactions on Image Processing*, **7**(1), 27–41, 1998.

[38] P. Thévenaz, T. Blu, and M. Unser, Interpolation Revisited, *IEEE Transactions on Medical Imaging*, **19**(7), 739–758, 2000.

[39] P. Thévenaz and M. Unser, Optimization of Mutual Information for Multiresolution Image Registration, *IEEE Transactions on Image Processing*, **9**(12), 2083–2099, 2000.

[40] M. Unser, Splines: A Perfect Fit for Signal and Image Processing, *IEEE Signal Processing Magazine*, **16**(6), 22–38, 1999.

[41] J. West, J.M. Fitzpatrick, M.Y. Wang, B.M. Dawant, C.R. Maurer Jr., R.M. Kessler, R.J. Maciunas, C. Barillot, D. Lemoine, A. Collignon, F. Maes, P. Suetens, D. Vandermeulen, P.A. van den Elsen, P.F. Hemler, S. Napel, T.S. Sumanaweera, B. Harkness, D.L.G. Hill, C. Studholme, G. Malandain, X. Pennec, M.E. Noz, G.Q. Maguire Jr., M. Pollack, C.A. Pelizzari, R.A. Robb, D. Hanson, and R.P. Woods, Comparison and Evaluation of Retrospective Intermodality Image Registration Techniques, in *Proceedings of the SPIE International Conference on Image Processing: Medical Imaging*, K.M. Hanson, Ed., Newport Beach CA, USA, February 12-15, 1996, SPIE, **2710**, 332–347.

[42] J. West, J.M. Fitzpatrick, M.Y. Wang, B.M. Dawant, C.R. Maurer Jr., R.M. Kessler, R.J. Maciunas, C. Barillot, D. Lemoine, A. Collignon, F. Maes, P. Suetens, D. Vandermeulen, P.A. van den Elsen, S. Napel, T.S. Sumanaweera, B. Harkness, P.F. Hemler, D.L.G. Hill, D.J. Hawkes, C. Studholme, J.B.A. Maintz, M.A. Viergever, G. Malandin, X. Pennec, M.E. Noz, G.Q. Maguire Jr., M. Pollack, C.A. Pelizzari, R.A. Robb, D. Hanson, and R.P. Woods, Comparison and Evaluation of Retrospective Intermodality Brain Image Registration Techniques, *Journal of Computer Assisted Tomography*, **21**(4), 554–566, 1997.

[43] J. West, J.M. Fitzpatrick, M.Y. Wang, B.M. Dawant, C.R. Maurer Jr., R.M. Kessler, and R.J. Maciunas, Retrospective Intermodality Registration Techniques for Images of the Head: Surface-Based Versus Volume-Based, *IEEE Transactions on Medical Imaging*, **18**(2), 144–150, 1999.

[44] H. Weyl, Über die Gleichverteilung von Zahlen mod Eins, *Annals of Mathematics*, **77**, 313–352, 1916.

[45] R.P. Woods, J.C. Mazziotta, and S.R. Cherry, MRI-PET Registration with Automated Algorithm, *Journal of Computer Assisted Tomography*, **17**(4), 536–546, 1993.