

# Fast Computation of Mutual Information in a Variational Image Registration Approach

Stefan Heldmann, Oliver Mahnke, Daniel Potts,  
Jan Modersitzki and Bernd Fischer

Institute of Mathematics, University of Lübeck,  
Wallstraße 40, 23560 Lübeck, Germany  
E-Mail: {heldmann,mahnke,potts,modersitzki,fischer}@math.uni-luebeck.de

**Abstract.** In this paper we present a novel method for computing the Mutual Information of images using recently developed non-equidistant fast Fourier-transform (NFFT) techniques. Standard approaches suffer from the problem that some sort of quantization is needed in order to apply a fast equidistant FFT-technique. For the new method no quantization is necessary which on one hand leads to an improved registration accuracy and on the other hand to a straight forward implementation. The evaluation is done for MR brain registration as well as for synthetic examples.

## 1 Introduction

Image registration is one of today's most challenging problems in digital imaging. Given a reference image and a template image the task is to find a geometric transformation that maps the template onto the reference, such that the images are similar and corresponding points match. Particularly in medical imaging the difficulty arises that images are taken from several different devices, like e.g. Computer Tomography, Magnetic Resonance Imaging, or Ultra Sound scanners. Thus, their intensities cannot be taken directly to measure the images similarity. Recent studies show that maximizing the Mutual Information of the images performs very successful for image registration in a multimodal situation [1,2,3,4]. In order to compute the Mutual Information one has to estimate in one way or another the intensity distribution of the given images. This is a tricky problem for discrete digital images. However, the accuracy of this estimation plays a very important role for the outcome of the registration as well as for the needed computational effort.

Basically, one has to distinguish between two approaches for estimating the distribution function. The first one does construct a discrete approximation whereas the second one deals with continuous intensity distributions. Methods for constructing discrete distributions are histogram based and therefore are easy to compute. However, their accuracy is limited by the quantization of the intensities. Furthermore, since efficient numerical optimization schemes in general make use of the derivative of the distribution function, they cannot be applied directly

in the discrete setting. Here, the needed derivative has to be approximated which may slow down the convergence of the overall numerical scheme. For constructing continuous distribution functions, typically a so-called PARZEN-window estimator is used. Such an estimator is the sum of PARZEN-window-functions, usually Gaussians that are centered at points of a drawn sample. Its computational complexity is much higher than the one for computing a histogram based estimate and is directly connected to the sample size. To reduce the computational effort, the PARZEN-estimator is often constructed from a small number of samples [1]. Its derivative, however, is always known analytically.

In this note we propose a new method for estimating the continuous distribution function. It is based on special band-limited PARZEN-window-functions, that are approximations of conventional window functions, such as Gaussians or Cauchy densities. The main point is, that these functions enables one to employ recently developed fast Fourier transforms at non-equispaced knots (NFFTS) [5,6]. Using this approach, the computational complexity drops drastically. More precisely, the computational cost to approximate the PARZEN-estimator constructed from  $M$  samples at  $N$  points reduces from  $\mathcal{O}(NM)$  to  $\mathcal{O}(N + M)$  operations. Thus, we are able to use very large samples, yielding a reliable and robust density estimate.

## 2 A variational approach for image registration

Given two  $d$ -dimensional images  $R, T : \mathbb{R}^d \rightarrow \mathbb{R}$ , one is interested to find a displacement  $u : \mathbb{R}^d \rightarrow \mathbb{R}^d$  that maps the template  $T$  onto the reference  $R$ , such that  $T \circ (I - u)$  is similar to  $R$  on a domain  $\Omega \subset \mathbb{R}^d$ . A common approach is to minimize the joint-functional (see, for example [7])

$$\mathcal{J}[R, T; u] := [R, T; u] + \alpha \mathcal{S}[u], \quad \alpha > 0, \quad (1)$$

with a distance measure and a so-called smoother  $\mathcal{S}$ . The distance measure rates the similarity of  $R$  and the deformed template  $T_u := T \circ (I - u)$ . The smoother  $\mathcal{S}$  penalizes unwanted deformations. The parameter  $\alpha$  weights the similarity of the images versus the smoothness of the displacement.

**Mutual information as distance measure.** We use Mutual Information to measure the distance of images. Here, one does regard the two images as random variables  $R, T_u : \Omega \rightarrow \mathbb{R}$  with the joint intensity density  $p^{R, T_u} : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  and marginal densities  $p^R, p^{T_u} : \mathbb{R} \rightarrow \mathbb{R}$ . The Mutual Information measures the similarity of  $R$  and  $T_u$  by the KULLBACK-LEIBLER-distance. Then the wanted distance measure is given by [3,4]

$$\text{MI}[R, T; u] := -\text{MI}[R, T_u] = - \int_{\mathbb{R}^2} p^{R, T_u}(r, t) \log \frac{p^{R, T_u}(r, t)}{p^R(r)p^{T_u}(t)} d(r, t). \quad (2)$$

**The smoother.** Several smoothers have been proposed in literature. Among those is the so-called curvature smoother

$$\mathcal{S}^{\text{CURV}}[u] := \sum_{\ell=1}^d \int_{\Omega} (\Delta u_{\ell}(x))^2 dx, \quad (3)$$

introduced by FISCHER&MODERSITZKI [7], which is used here.

**The optimization method.** To compute a solution of the minimization problem (1) we make use of the fact, that the first variation of the joint functional  $\mathcal{J}$  has to vanish for a minimizer, which holds if and only if [7,8]

$$\alpha \Delta^2 u + f(u) = 0 \quad \text{on } \Omega, \quad \nabla u_\ell = \nabla \Delta u_\ell = 0 \quad \text{on } \partial\Omega, \ell = 1, 2, \dots, d, \quad (4)$$

with the so-called *forces*

$$f(x, u(x)) := L^{R, Tu}(R(x), T(x - u(x))) \cdot \nabla T(x - u(x)), \quad (5)$$

whereby  $L^{R, Tu}(r, t) := \frac{\partial_t p^{Tu}(t)}{p^{Tu}(t)} - \frac{\partial_t p^{R, Tu}(r, t)}{p^{R, Tu}(r, t)}$ . Thus a minimizer is a solution of the boundary value problem (4).

**The numerical method.** To compute a solution of the boundary value problem (4) we apply a semi-discrete time-marching method. To approximate spatial derivatives we use finite difference. This leads to a linear system that can be efficiently solved within  $\mathcal{O}(N \log N)$  operations [7].

### 3 Efficient PARZEN-window estimation

To evaluate the forces (5) we have to estimate the densities  $p^{R, T}$  and  $p^T$ . Therefore we use a PARZEN-estimator as proposed in [1]:

$$\hat{p}^X(x) = \frac{1}{M} \sum_{j=1}^M K(x - X_j), \quad (6)$$

whereby  $X_1, X_2, \dots, X_M$  denote realizations of the random variable  $X$  and  $K$  denotes the so-called PARZEN-window. As window function we use a band-limited approximation of the Gaussian

$$K(x) := \sum_{k=-n/2}^{n/2-1} \alpha_k \exp(i2\pi kx), \quad \alpha_k := \exp(-2\sigma^2 \pi^2 (k/n)^2), \quad (7)$$

where  $\alpha_k$  is the Fourier-coefficient of a Gaussian with the standard deviation  $\sigma$  at the frequency  $k/n$ . From (6) and (7) we obtain

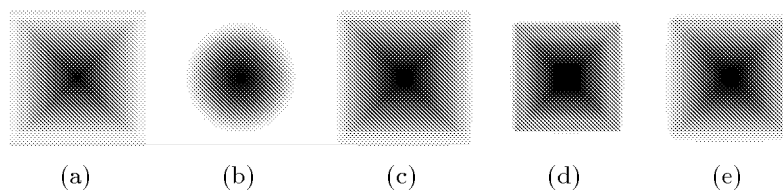
$$\hat{p}^X(x) = \frac{1}{M} \sum_{k=-n/2}^{n/2-1} \alpha_k \underbrace{\left( \sum_{j=1}^M \exp(-i2\pi kX_j) \right)}_{\text{NFFT}^T} \exp(i2\pi kx). \quad (8)$$

Using the non-equidistant fast Fourier transform NFFT and its transposed version NFFT<sup>T</sup> [5] we are able to evaluate  $\hat{p}^X(x_\ell)$ ,  $\ell = 1, 2, \dots, N$  with  $\mathcal{O}(n \log n + M + N)$  operations, where  $n$  is a constant depending on the approximation. The derivative  $\frac{d}{dx} \hat{p}^X$  can be computed along the same lines. In contrast to other fast algorithms like the discrete Gauss transform [9], our method is easy to implement and can be adapted to other window functions, e.g. Cauchy densities or B-splines, by simply choosing different coefficients  $\alpha_k$ .

**Table 1.** Timings to evaluate the forces (5) for the NFFT based and the histogram based method with  $b$  the number of bins (AMD Athlon XP 2700+, SuSE Linux 8.2).

$N, M$	NFFT	$b = 32$	$b = 64$	$b = 128$	$b = 256$	$b = 512$	$b = 1024$
$128^2$	0.30s	0.03s	0.04s	0.14s	0.77s	3.01s	12.11s
$256^2$	0.91s	0.11s	0.12s	0.22s	0.85s	3.10s	12.20s
$512^2$	3.41s	0.57s	0.58s	0.68s	1.31s	3.56s	12.65s
$1024^2$	13.34s	2.27s	2.29s	2.38s	3.02s	5.29s	14.36s

**Fig. 1.** Registration results; (a) reference; (b) template; (c) NFFT based method; (d),(e) histogram based method (128 and 256 bins).

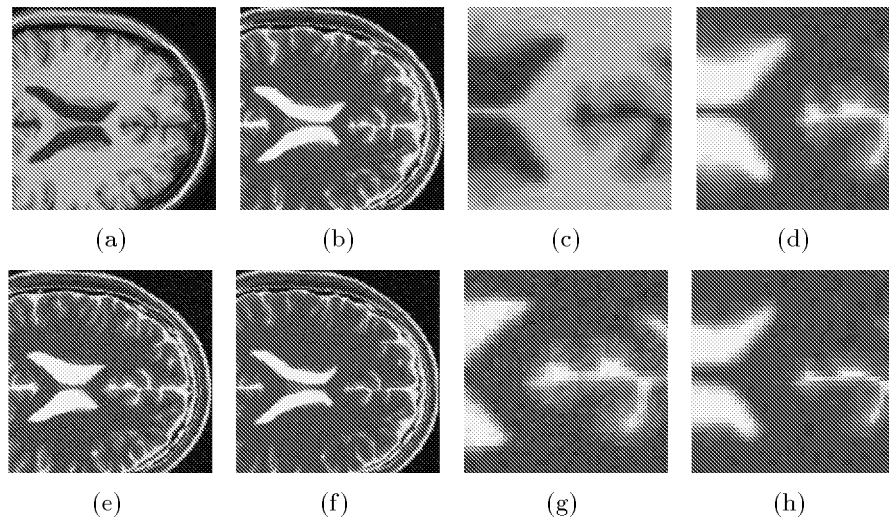


## 4 Results

In this section we compare our new method with a histogram based method. To this end, we compute the histogram of a random sample and subsequently smooth it by convolving it with a discrete Gaussian. The needed derivative is approximated by standard finite differences. Finally, we evaluate the obtained estimates by a bilinear interpolation. Using a sample with  $M$  points and a histogram with  $b$  bins, applying a FFT technique the computational complexity of this method is  $\mathcal{O}(b^2 \log b + M + N)$ , with  $N$  the number of pixels (cf. tab. 1). In contrast to the naive implementation, the NFFT and histogram based methods reduce the complexity from  $\mathcal{O}(NM)$  to  $\mathcal{O}(N + M)$ , both. The binning of intensities leads to inaccurate results for histogram based methods in areas with low local contrast, because small gray-value changes are not taken into account if the bin size is chosen too small. The more bins used, the more accurate is the outcome of the registration. Since our method allows for non-quantized sample values it is not affected by the *binning effect*. Fig. 1 shows the registration of low contrast images with intensities in  $[0, 25]$ . To illustrate the binning effect, the intensity range  $[0, 255]$  was discretized with 128 and 256 bins. Our method leads to a nearly perfect matching (c.f. fig. 1(c)), whereas the histogram-based methods fail to produce an accurate result (c.f. figs. 1(d),(e)).

In a second experiment we registered simulated T1/T2 brain images from the brain-web database. As expected, the histogram based method leads to misregistration of regions with low local contrast, e.g. the border between gray and white matter (c.f. figs. 2(h),(d)).

**Fig. 2.** Registration results; (a) reference; (e) template; (b) NFFT based; (f) histogram based (32 bins); (c) reference (detail); (g) template (detail); (d) NFFT based (detail); (h) histogram based method (detail).



## References

1. P. Viola. *Alignment by Maximization of Mutual Information*. PhD thesis, MIT, 1995.
2. A. Collignon, F. Maes, D. Vandermeulen, P. Suetens, and G. Marchal. *Automated multi-modality image Registration based on information theory*. Information Processing in Medical Imaging, 1995.
3. G. Hermosillo. *Variational Methods for Multimodal Image Registration*. PhD thesis, Université de Nice, France, 2002.
4. E. D'Agostino, F. Maes, D. Vandermeulen, P. Suetens. *A viscous flow model for multimodal non-rigid image registration using mutual information*. MICCAI 2002, Tokyo, 23-26 September, 2002.
5. S. Kunis, D. Potts. *Software package, C subroutine library*. <http://www.math.uni-luebeck.de/potts/nfft>, 2003
6. D. Potts, G. Steidl. *Fast Summation at non-equidistant knots by NFFTs*. SIAM J. Sci. Comput., 24:2013-2037, 2003.
7. B. Fischer, J. Modersitzki. *Curvature based image registration*. JMIV 18(1), 2003.
8. O. Mahnke, S. Heldmann, J. Modersitzki, B. Fischer. *A variational approach for maximizing mutual information*. In preparation, 2004.
9. L. Greengard and X. Sun. *The fast Gauss transform*. Doc. Math. J. DMV, 3:575-584, 1998.