Generalized Partial Volume

an inferior density estimator to Parzen Windows for normalized mutual information

Darkner, Sune; Sporring, Jon

Published in:
Information Processing in Medical Imaging

DOI:
10.1007/978-3-642-22092-0_36

Publication date:
2011

Document version
Peer reviewed version

Generalized Partial Volume: An Inferior Density Estimator to Parzen Windows for Normalized Mutual Information

Sune Darkner and Jon Sporring

eScience Center, Department of Computer Science, University Of Copenhagen
Universitetsparken 1, DK-2100 Copenhagen, Denmark
{darkner, sporring}@diku.dk

Abstract. Mutual Information (MI) and normalized mutual information (NMI) are popular choices as similarity measure for multimodal image registration. Presently, one of two approaches is often used for estimating these measures: The Parzen Window (PW) and the Generalized Partial Volume (GPV). Their theoretical relation has so far been unexplored. We present the direct connection between PW and GPV for NMI in the case of rigid and non-rigid image registration. Through step-by-step derivations of PW and GPV we clarify the difference and show that GPV is algorithmically inferior to PW from a model point of view as well as w.r.t. computational complexity. Finally, we present algorithms for both approaches for NMI which is comparable in speed to Sum of Squared Differences (SSD), and we illustrate the differences between PW and GPV on a number of registration examples.

Keywords: Similarity measure, registration , normalized mutual information, density estimation, scale space, locally orderless images.

1 Introduction

Mutual information (MI) and its normalized version (NMI) are considered state of the art for image registration. MI and NMI are particularly useful for registering Magnetic Resonance Images (MRI) to MRI as well as for multimodal image registration in general. MI and NMI are entropy based measures and hence rely on intensity distributions. Intensity distributions are most often approximated by discrete histograms, which poses a challenge to gradient based methods. Today, one of two approaches is often used for estimating these measures: The Parzen Window (PW) [21] and the Generalized Partial Volume (GPV) [3, 10]. Empirical comparisons have previously been presented, but their theoretical connection has so far been unexplored despite the fact that both are used in the same context. We present the derivations of both PW and GPV in a joint theoretical context. This gives novel insight to the relation between PW and GPV enabling a theoretical evaluation and comparison of the two. We show that PW and GPV are special cases of histogram registration using Locally Orderless

Images (LOI) [7] on NMI. The concept of LOI allow us to treat derivatives in measurement, integration, and intensity space in a well-posed manner, as well as offer a scale-space formulation of these spaces. From this we formulate algorithms, analytically compare their speed, and discuss the different approaches.

We evaluate PW and GPV in the setting of LOI by a series of simple experiments with numerically the exact same prerequisites, interpolation, regularization, and optimization on publicly available data: 3D cardiac MRI [1], and 3D T1 brain MRI [12] for rigid intra- and inter-patient registration as well as non-rigid intra patient registration. For each registration we report standard deviation and mean as a function of method parameters, and we conclude that there are differences between the proposed algorithms, and differences reported originates from approximations and smoothing in different spaces: GPV in namely the isophote domain and in PW the image and intensity domain. This leads us to conclude that PW is more attractive and in general superior density estimator to GPV for MI and NMI.

1.1 Previous work

The use of Mutual Information (MI) for image registration was originally proposed by [4, 21]. An extensive overview was given in [17]. Normalized Mutual Information (NMI) was introduced as a more robust alternative especially designed for multi modal image registration [19]. The first implementations relied on Powell's method [10], hill climbing [19], or similar methods without gradients, which were accurate but slow. A GPU speedup was suggested in [13]. Today, state-of-the-art implementations are gradient based methods and group in two algorithm types: The first type is based on Parzen Windows (PW) [21] and relies on the fact that the marginal and the joint histograms are made continuous by using different kernels, e.g., Gaussian or B-splines [20]. The second type is based on Generalized Partial Volume (GPV), where the distribution is sampled from the image directly [10]. Analytical derivatives of this method were presented in [11] and a generalization using B-splines was presented in [3]. A variational method relating to LOI [7] for MI (and other measures) was presented in [6]. GPV and PW was compared numerically in [8] concluding that PW is precise and GPV has a larger convergence radius. MI and NMI are notorious for their local minima and difficulty of implementation and the choice of interpolation scheme greatly influence the smoothness of the objective function. Some investigations into this can be found in [5, 16]. An alternative approach is the Conditional Mutual Information [9].

In this article, we investigate PW and GPV for NMI, using differential calculus in a thorough step-by-step presentation. The derivations may be seen as an alternative to the variational approach in [6], our's holds the same generality, but lead to much faster algorithms, since numerical issues become obvious, and our derivation lead to a direct comparison between PW and GVP. The remainder of this article is organized as follows: In Section 2 the general registration framework is described. In Section 3 we discuss LOI as a basis for analyzing GPV and PW as well as demonstrate their similarity and derive fast algorithms.

In Section 5 we compare the methods experimentally, and finally, in Section 6 we give our conclusions.

2 Image registration

Image registration is the process of transforming one image $\tilde{I} : \Omega \to \Gamma$, where $\Omega \subseteq \mathbb{R}^N$ and $\Gamma \subseteq \mathbb{R}$, w.r.t. a reference image $R : \Omega \to \Gamma$ such that some functional $\mathcal{F}(\tilde{I}, R)$ is minimized. We consider diffeomorphic transformation of NM parameters, $\boldsymbol{\phi} : \Omega \times \mathbb{R}^{NM} \to \Omega$, and for brevity we write $I = \tilde{I} \circ \boldsymbol{\phi}$. The general form of $\mathcal{F}$ is,

$$\mathcal{F} = \mathcal{M}(I, R) + \mathcal{S}(\boldsymbol{\phi}), \tag{1}$$

where $\mathcal{M}$ is a (dis)similarity measure between the images and $\mathcal{S}(\boldsymbol{\phi})$ is a regularization term. Regularization is almost always required in order to obtain a unique solution, since image registration is generally ill-posed [14]. We use Riemannian Elasticity [15],

$$\mathcal{S}(\boldsymbol{\phi}) = \frac{\mu}{4} \sum_i \log^2 \epsilon_i + \frac{\lambda}{8} \left(\sum_i \log \epsilon_i \right)^2, \tag{2}$$

where ϵ_i is the i'th squared eigenvalue of the Jacobian of the transformation ϕ, μ and λ are regularization parameters.

The focus of this paper is the similarity measure, and we consider NMI for $\mathcal{M}$ [19],

$$\mathcal{M}_{\text{NMI}} = \frac{\mathcal{H}_I + \mathcal{H}_R}{\mathcal{H}_{I,R}}, \tag{3}$$

where $\mathcal{H}$ denotes the marginal and the joint entropy of the intensity distribution [18], specifically,

$$\mathcal{H}_I = -\int_\Gamma p_I(i) \log p_I(i)\, di, \qquad \mathcal{H}_R = -\int_\Gamma p_R(i) \log p_R(i)\, di, \tag{4}$$

$$\mathcal{H}_{I,R} = -\int_{\Gamma^2} p_{I,R}(i,j) \log p_{I,R}(i,j)\, di \wedge dj, \tag{5}$$

using the natural logarithm for convenience, and the intensity and joint intensity distributions, $p(\cdot) : \Gamma \to \mathbb{R}_+$ and $p(\cdot,\cdot) : \Gamma^2 \to \mathbb{R}_+$, are estimated by the histogram and joint histogram of the intensity values. NMI has proven to be very powerful for registration of medical images in general.

Finally, we consider Uniform B-splines [2] for coordinate transformation,

$$\boldsymbol{x} = \boldsymbol{\phi}(\tilde{\boldsymbol{x}}, \boldsymbol{\Phi}) = \sum_{m=0}^{M} \boldsymbol{\phi}_m \prod_{n=1}^{N} f_{m,t}(\tilde{x}_n), \tag{6}$$

where $\boldsymbol{\Phi} = [\boldsymbol{\phi}(\tilde{\boldsymbol{x}}_1), \ldots, \boldsymbol{\phi}(\tilde{\boldsymbol{x}}_M)] \in \Omega^M$ is a matrix of values of the transformation at regular grid coordinates $\tilde{\boldsymbol{x}}_m$, $\tilde{\boldsymbol{x}} = [\tilde{x}_1, \ldots, \tilde{x}_N]^T$ is the evaluation point, and where the coordinate wise interpolation function, $f_{m,t} : \mathbb{R} \to \mathbb{R}$, is given by the Cox-de Boor recursion formula [2].

In the following, we will study the implication of the 3 independent scales of the local histogram on the registration problem, and we derive the PW and GPV for NMI in the context of LOI for usage in a quasi-Newton method.

3 Locally orderless images (LOI)

To fully understand the connection between PW and GPV we first shortly review the concept of locally orderless images (LOI), the theoretical foundation of histogram generation.

The critical element for NMI based registration methods is the generation of the density distribution and efficiently calculating the gradient of (1), especially the gradient of (3), and in parts the gradient of $\boldsymbol{\phi}$. We study NMI registration through LOI [7], which extends the concept of histograms with the 3 fundamental scales, the amount of spatial smoothing of the images (image smoothing), the amount of histogram smoothing (intensity smoothing), and the size of the window (the partial volume) for calculating local histograms. Thus, a local histogram is written as,

$$h_I(i, \boldsymbol{x}, \boldsymbol{\Phi}, \alpha, \beta, \sigma) = P(I(\boldsymbol{x}, \boldsymbol{\Phi}, \sigma) - i, \beta) * W(\boldsymbol{x}, \alpha), \tag{7}$$

$$I(\boldsymbol{\psi}, \boldsymbol{\Phi}, \sigma) = I(\boldsymbol{x}) * K(\boldsymbol{x}, \sigma), \tag{8}$$

where $d\psi_N$, $i \in \Gamma$, P is a Parzen window of intensity or tonal scale $\beta \in \mathbb{R}_+$, K is a spatial measurement kernel of scale $\sigma \in \mathbb{R}_+$, W is an integration window of scale $\alpha \in \mathbb{R}_+$ and located at $\boldsymbol{x}$, $\cdot * \cdot$ is the convolution operator taken w.r.t. the variable $\boldsymbol{x}$, and $\boldsymbol{\Phi}$ denotes the parameters for the transformation. The histogram h_R is defined similarly independently of $\boldsymbol{\Phi}$. In [7] it is proposed to use $P(i, \beta) = e^{-i^2/(2\beta^2)}$, and $K(\boldsymbol{x}, \sigma) = W(\boldsymbol{x}, \sigma) = e^{-\boldsymbol{x}^T\boldsymbol{x}/(2\sigma^2)}/(2\pi\sigma^2)^{N/2}$, and this structure is called the Locally Orderless Image. The distributions are obtained by normalizing to unity,

$$p_I(i|\boldsymbol{x}, \boldsymbol{\Phi}, \alpha, \beta, \sigma) \simeq \frac{h_I(i, \boldsymbol{x}, \boldsymbol{\Phi}, \alpha, \beta, \sigma)}{\int_\Gamma h_I(j, \boldsymbol{x}, \boldsymbol{\Phi}, \alpha, \beta, \sigma) dj}, \tag{9}$$

$$p_I(i|\boldsymbol{\Phi}, \alpha, \beta, \sigma) = \frac{1}{|\Omega|} \int_\Omega p_I(i|\boldsymbol{x}, \boldsymbol{\Phi}, \alpha, \beta, \sigma)\, d\boldsymbol{x}, \tag{10}$$

where we have assumed (conditional) independence and uniformity such that $p_I(i, \boldsymbol{x}|\boldsymbol{\Phi}, \alpha, \beta, \sigma) = p_I(i|\boldsymbol{x}, \boldsymbol{\Phi}, \alpha, \beta, \sigma)/|\Omega|$. The density p_R is defined in a similar

manner. As [6], we extend the concept to the joint distributions as follows,

$$h_{I,R}(i,j,\boldsymbol{x},\boldsymbol{\Phi},\alpha,\beta,\sigma) = (P(I(\boldsymbol{x},\boldsymbol{\Phi},\sigma)-i,\beta)P(J(\boldsymbol{x},\sigma)-j,\beta)) * W(\boldsymbol{x},\alpha), \tag{11}$$

$$p_{I,R}(i,j|\boldsymbol{x},\boldsymbol{\Phi},\alpha,\beta,\sigma) \simeq \frac{h_{I,R}(i,j,\boldsymbol{\Phi},\boldsymbol{x},\alpha,\beta,\sigma)}{\int_{\Gamma^2} h_{I,R}(k,l,\boldsymbol{x},\alpha,\beta,\sigma)\,dk\wedge dl}, \tag{12}$$

$$p_{I,R}(i,j|\boldsymbol{\Phi},\alpha,\beta,\sigma) = \frac{1}{|\Omega|}\int_{\Omega} p_{I,R}(i,j|\boldsymbol{\Phi},\boldsymbol{x},\alpha,\beta,\sigma)\,d\boldsymbol{x}, \tag{13}$$

where we also have assumed (conditional) independence and uniformity such that $p_{I,R}(i,j,\boldsymbol{x}|\boldsymbol{\Phi},\alpha,\beta,\sigma) = p_{I,R}(i,j|\boldsymbol{x},\boldsymbol{\Phi},\alpha,\beta,\sigma)/|\Omega|$.

3.1 First Order Structure

In order to use quasi-Newton methods for optimization, we need to derive the gradient of (1) w.r.t. the parameters of the uniform cubic b-spline, $\boldsymbol{\Phi}$. We use the notation of differentials, $dg(x) = Dg(x)\,dx$, where D is the partial derivative operator. Note that $d\boldsymbol{x}$ is a vector of differentials, not the wedge product of its elements, when used in relation to differentiation. Further, we will only write up non-zero terms that depend on $d\boldsymbol{\Phi}$. The differential of (1) is,

$$d\mathcal{F} = d\mathcal{M} + d\mathcal{S}, \tag{14}$$

where arguments have been omitted for brevity. Ignoring the regularization term we focus on the differential of the similarity measure. The differential of (3) is,

$$d\mathcal{M}_{\text{NMI}} = \frac{(d\mathcal{H}_I + d\mathcal{H}_R)\mathcal{H}_{I,R} - (\mathcal{H}_I + \mathcal{H}_R)d\mathcal{H}_{I,R}}{\mathcal{H}_{I,R}^2}. \tag{15}$$

The entropy, $\mathcal{H}_R$, is independent of $\boldsymbol{\phi}$, hence $d\mathcal{H}_R = 0$. Further,

$$d\mathcal{H}_I = -\int_{\Gamma} dp_I\,(\log p_I + 1)\,di, \tag{16}$$

$$d\mathcal{H}_{I,R} = -\int_{\Gamma^2} dp_{I,R}(\log p_{I,R} + 1)\,di\wedge dj. \tag{17}$$

Using Leibniz integration rule, the differentials of the distributions are given as

$$dp_I(i,\boldsymbol{\Phi}) = \frac{1}{|\Omega|}\int_{\Omega} dp_I(i|\boldsymbol{x},\boldsymbol{\Phi})\,d\boldsymbol{x}, \tag{18}$$

$$dp_I(i|\boldsymbol{x},\boldsymbol{\Phi}) \simeq \frac{dh_I(i,\boldsymbol{x},\boldsymbol{\Phi})}{\int_{\Gamma} h_I(j,\boldsymbol{x},\boldsymbol{\Phi})dj} - \frac{h_I(i,\boldsymbol{x},\boldsymbol{\Phi})\int_{\Gamma} dh_I(j,\boldsymbol{x},\boldsymbol{\Phi})dj}{\left(\int_{\Gamma} h_I(j,\boldsymbol{x},\boldsymbol{\Phi})dj\right)^2}, \tag{19}$$

$$dh_I(i,\boldsymbol{x},\boldsymbol{\Phi}) = \left(dP(I(\boldsymbol{x},\boldsymbol{\Phi},\sigma)-i,\beta) * W(\boldsymbol{x},\alpha)\right), \tag{20}$$

where irrelevant arguments have been omitted for brevity. Likewise, we have:

$$dp_{I,R}(i,j) = \frac{1}{|\Omega|}\int_\Omega dp_{I,R}(i,j|\boldsymbol{x})\,d\boldsymbol{x}, \tag{21}$$

$$dp_{I,R}(i,j|\boldsymbol{x}) \simeq \frac{dh_{I,R}(i,j,\boldsymbol{x})}{\int_{\Gamma^2} h_{I,R}(k,l,\boldsymbol{x})\,dk\wedge dl} - \frac{h_{I,R}(i,j,\boldsymbol{x})\int_{\Gamma^2} dh_{I,R}(k,l,\boldsymbol{x})\,dk\wedge dl}{\left(\int_{\Gamma^2} h_{I,R}(k,l,\boldsymbol{x})\,dk\wedge dl\right)^2}, \tag{22}$$

$$dh_{I,R}(i,j,\boldsymbol{x}) = \left(dP(I(\boldsymbol{\psi},\boldsymbol{\Phi},\sigma)-i,\beta)\,P(J(\boldsymbol{\psi},\sigma)-j,\beta)\right) * W(\boldsymbol{x}-\boldsymbol{\psi},\alpha). \tag{23}$$

3.2 The Parzen window (PW)

Originally MI was proposed in [21] using PW as density estimator. In the following we will examine this special case of LOI, often used in the literature. Consider (7) and let $\alpha\to\infty$. In that case, the window h_I simplifies as,

$$h_I(i,\boldsymbol{x},\boldsymbol{\Phi},\alpha,\beta,\sigma) \to \text{const.}\int_\Omega P(I(\boldsymbol{\psi},\boldsymbol{\Phi},\sigma)-i,\beta)\,d\boldsymbol{\psi}, \tag{24}$$

$$p_I(i|\boldsymbol{\Phi},\alpha,\beta,\sigma) \to \frac{\int_\Omega P(I(\boldsymbol{\psi},\boldsymbol{\Phi},\sigma)-i,\beta)\,d\boldsymbol{\psi}}{\int_\Gamma\int_\Omega P(I(\boldsymbol{\psi},\boldsymbol{\Phi},\sigma)-j,\beta)\,d\boldsymbol{\psi}\wedge dj}. \tag{25}$$

Choosing

$$P(i,\beta) = e^{-i^2/(2\beta^2)}, \tag{26}$$

we find that

$$\int_\Gamma\int_\Omega P(I(\boldsymbol{\psi},\boldsymbol{\Phi},\sigma)-j,\beta)\,d\boldsymbol{\psi}\wedge dj = |\Omega|\sqrt{2\pi\beta^2}, \tag{27}$$

and

$$p_I(i|\boldsymbol{\Phi},\alpha,\beta,\sigma) \to \frac{1}{|\Omega|\sqrt{2\pi\beta^2}}\int_\Omega e^{\frac{-(I(\boldsymbol{x},\boldsymbol{\Phi},\sigma)-i)^2}{2\beta^2}}\,d\boldsymbol{x}. \tag{28}$$

Likewise, we have

$$p_{I,R}(i,j|\boldsymbol{\Phi},\alpha,\beta,\sigma) \to \frac{1}{|\Omega|2\pi\beta^2}\int_\Omega e^{-\frac{(I(\boldsymbol{x},\boldsymbol{\Phi},\sigma)-i)^2+(R(\boldsymbol{x},\sigma)-j)^2}{2\beta^2}}\,d\boldsymbol{x}. \tag{29}$$

This is precisely the Parzen window method using a Gaussian kernel with infinite support [21]. Similar results are obtained for any integrable $P(i,\beta)$. The PW can be interpreted as a globally orderless image, as $\boldsymbol{W}$ defining the locality extends globally. Further, since both (28) and (29) obey the diffusion equation w.r.t. $\beta^2/2$, we may use Green's theorem and write,

$$p_I(i|\sqrt{\beta_0^2+\beta^2}) = p_I(i|\beta_0) * G(i,\beta), \tag{30}$$

$$p_{I,R}(i,j|\sqrt{\beta_0^2+\beta^2}) = p_{I,R}(i,j|\beta_0) * G([i,j]^T,\beta), \tag{31}$$

for fast computation of a range of Parzen window sizes.

3.3 Generalized Partial Volume (GPV)

Shortly after the introduction of PW partial volume (PV) was introduced in [10] and extended to GPV in [3]. In the context of LOI GPV can be derived as follows:

$$dh_I = d\left(P(I(\boldsymbol{x},\boldsymbol{\Phi},\sigma)-i,\beta) * W(\boldsymbol{x},\alpha)\right) \tag{32}$$
$$= P(I(\tilde{\boldsymbol{x}},\boldsymbol{\Phi},\sigma)-i,\beta) * \left(D_{\boldsymbol{x}} W(\boldsymbol{x},\alpha)\right), \tag{33}$$

For the joint histograms partial volume approximate as follows:

$$h_{I,R}(i,j,\boldsymbol{x},\alpha,\beta,\sigma)$$
$$= \int_{\Omega} P(I(\boldsymbol{\psi},\sigma)-i,\beta)P(J(\boldsymbol{\psi},\sigma)-j,\beta)W(\boldsymbol{x}-\boldsymbol{\psi},\alpha)\,d\boldsymbol{\psi} \tag{34}$$
$$\approx P(J(\boldsymbol{x},\sigma)-j,\beta)\int_{\Omega} P(I(\boldsymbol{\psi},\sigma)-i,\beta)W(\boldsymbol{x}-\boldsymbol{\psi},\alpha)\,d\boldsymbol{\psi} \tag{35}$$
$$= P(J(\boldsymbol{x},\sigma)-j,\beta)\left((P(I(\boldsymbol{\phi}(\tilde{\boldsymbol{x}}),\sigma)-i,\beta)) * W(\boldsymbol{x},\alpha)\right) \tag{36}$$

Hence, the differential w.r.t.x

$$dh_{I,R}(i,j,\boldsymbol{x},\alpha,\beta,\sigma)$$
$$= P(J(\boldsymbol{x},\sigma)-j,\beta)\left((P(I(\boldsymbol{\phi}(\tilde{\boldsymbol{x}}),\sigma)-i,\beta)) * \left(D_{\boldsymbol{x}} W(x,\alpha)\right)\right), \tag{37}$$

Set P $(P(I(\boldsymbol{\psi},\boldsymbol{\Phi},\sigma)-i,\beta)$ and $P(J(\boldsymbol{x},\sigma)-j,\beta))$ to a boxcar functions

$$P(I(\boldsymbol{\psi},\boldsymbol{\Phi},\sigma)-i,\beta) = \begin{cases} 1 & \text{if} \quad -\frac{\beta}{2} \le \tilde{I}(\boldsymbol{\psi},\boldsymbol{\Phi},\sigma)-i < \frac{\beta}{2}, \\ 0 & \text{otherwise,} \end{cases} \tag{38}$$

where β is chosen such that $I(\boldsymbol{\psi},\boldsymbol{\Phi},\sigma)$ is mapped into non-coinciding isophotes curves. The motivation for this is that all isophotes can be evaluated at $\tilde{x}$ simultaneously. This is the generalized partial volume (GPV) scheme when integrating over the entire domain $\boldsymbol{\Omega}$. Thus GPV is small local histograms integrated to form the globally orderless image as in the PW approach.

4 Fast Implementations

This section presents a complexity analysis of the compact pseudo code required to implement SSD, PW and GPV. The purpose is to show that these measures can be implemented as very fast algorithms with only slightly more computations than SSD. We compare the computational complexity to SSD using identically interpolation scheme. All kernels used in implementations are 3rd order uniform B-splines as well as boxcar functions to reduce computational complexity. The code assumes 3D images, the use cubic B-splines for all kernels and K bins in the histograms. The pseudo code can be seen in Figure 1 including the analysis of the computational complexity. We assume that today's processors have equal

```
# Given 2 images, I and R, and the determinant of the
# transformation, det, as a function of space,
# calculate PW and GPV for NMI and SSD, based on N
# image evaluation points, and K marginal and K^2 joint
# histogram bins. Flops are based on cubic splines

FOR N evaluation points
  calculate image spline coeff. (60 flops)
  IF(SSD || PW)
    calculate derivative of image spline coeff. (48 flops)
  FOR 64 combinations of image spline coeff.
    IF(SSD || PW)
      update image at evaluation point (4 flops)
      update image gradient at evaluation point (12 flops)
    IF(GPV)
      update histograms (4 flops)
  IF(SSD)
    update residual (2 flops)
  IF(PW)
    calculate histogram spline coeff. (20 flops)
    FOR 16 histogram spline coeff.
      update histograms (2 flops)
IF(PW || GPV)
  calculate NMI and derivative on histograms (9*K^2+6K flops)
  FOR N evaluation points
    IF(GPV)
      calculate derivative of image spline coeff. (48 flops)
      FOR 64 combinations of image spline coeff.
        update derivative of histogram (16 flops)
    IF(PW)
      FOR 16 histogram spline coeff.
        update derivative of histogram (9 flops)
  update derivatives (3 flops)
# Total flop usage:
#   SSD: 1134N flops
#   PW:  1331N +9K^2 +6K flops
#   GPV: 1383N +9K^2 +6K flops
```

Fig. 1. Pseudo code for SSD, PW and GPV.

processing time of, e.g., sum, log, sin etc. From the pseudo code in Figure 1 and the complexity we see that the Parzen window can be considerable larger without being inferior to GPV in performance. This is of cause a rough estimate and implementations may vary, but the amount of computations for NMI using either GPV or PW are comparable in computational complexity to SSD using B-splines. W.r.t. memory, GPV requires $192 \times N \times 8$ bytes of memory to obtain the speed, where the PW only requires $8 \times N \times 8$ bytes (on 64-bit, double precision).

5 Experiments

For ease of comparison we restrict ourselves primarily to rigid registration examples with a few non-rigid included as well. It has previously been shown that PW is superior to GPV in accuracy, and thus we quantitatively evaluate the difference in estimated parameters with the prerequisites listed in Section 1. As the results in Tables 1 and 2 show, there are more or less pronounced differences numerically between the two methods. In some of the cases both PW or GPV got stuck in a local minima often simultaneously but with different minimas. Such problem can be solved using scale space.

param	rot x	rot y	rot z	tx	ty x	tz
Between subject (Brain)						
mean	0.0236	-0.0420	-0.0195	2.1120	1.7486	-0.4984
std	0.0556	0.2689	0.0781	7.5876	3.3922	1.6482
Intra subject (Brain)						
mean	0.0005	0.0001	-0.0004	-0.1304	-0.0326	0.1685
std	0.0016	0.0032	0.0008	0.7604	0.0825	0.7033
Intra subject (Cardiac)						
mean	0.0005	-0.0002	-0.0007	0.0629	0.0473	0.0142
std	0.0006	0.0005	0.0009	0.0883	0.0729	0.0406

Table 1. : Inter and intra subject rigid registration difference and standard deviation in pixels for T1-weighted MRI brain data and cardiac CINE-MRI data. As seen the two methods produces significantly different results especially between subjects. This originates from the smoothing in different domains and the use of the Hölder inequality. The differences are measured in voxels and radians.

Intra subjects non-rigid (Brain)	Difference
mean	-0.0182
std	2.446

Table 2. The difference in parameter estimates over 1800 parameters in 18 non-rigid intra patient T1-weighted MRI registrations in voxels. The cause of difference is the same as for rigid.

6 Discussion and Conclusion

The differences between the PW and GPV can be explained from a theoretical point of view: From (29) it is clear that GPV and PW smooth in 2 different spaces, thus the comparisons are interesting computation wise but less meaningful w.r.t. accuracy. Nevertheless, the approximation made by GPV substantiate that PW as reported in [8] is more accurate than GPV. In addition The Hölder inequality also accounts for some of the difference in value of the functional as

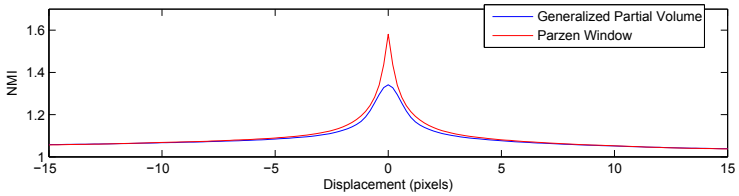

Fig. 2. The difference in NMI estimate for PW and GPW evaluated for 1 dimensional translational displacement. The observed difference originates from the smoothing in different domains and the use of the Hölder inequality.

does the choice of Parzen window. Our results also indicate that smoothing in the isophote domain $\boldsymbol{W}$ influences the objective function more than smoothing in the image domain I. The exact effects of smoothing in the 3 domains is left for future investigations.

From (**??**) we see that the Jacobian locally scales the integration compensating for the deformation. In some schemes this is omitted, reformulating the transformation to a re-sampling. The choice of approach should be based on the specific application; however, for density distributions the Jacobian should be present. The local scaling can cause instability in non-rigid, non-regularized settings with entropy based measures, as it tends to concentrate the mass at a single bin in the histogram, a global optimum for both MI and NMI. This effect is particularly pronounced in GPV, whereas the PW diffuses this effect over several bins in the $\boldsymbol{R}$ direction and not only the $\boldsymbol{I}$ direction. Some schemes overcome the missing Jacobian implicitly by composing small deformations and re-sampling.

To conclude, PW operates directly on intensity distributions and is therefore the natural choice for information theoretical measures such as NMI. In this paper we have shown that GPV and PW are special cases of LOI extended to joint probability distributions. From the perspective of LOI we have derived the PW and GPV, and shown how GPV makes a series of approximations and use of some special kernels with locality assumptions to achieve its low computational cost. The PW only rely on transforming LOI to globally orderless images, whereas the GPV rely approximations using the Hölder inequality to achieve computability, the use of a boxcar window as Parzen estimator and no image interpolation as a part of the model. The global entropy comes from integration over the entire domain. The approximations made in GPV along with smoothing in different spaces account for the difference in functional and in value of similarity measure makes PW far more attractive theoretically as well as in practice. In addition we have shown that PW is faster than the GPV and we substantiated the result from [8] that PW is more accurate. We therefore conclude that PW is a superior estimator for NMI compared to GPV.

References

1. Andreopoulos, A., Tsotsos, J.: Efficient and generalizable statistical models of shape and appearance for analysis of cardiac MRI. Medical Image Analysis 12(3), 335–357 (2008)
2. Boor, C.d.: A Practical Guide to Splines. Springer-Verlag (1978)
3. Chen, H., Varshney, P.: Mutual information-based CT-MR brain image registration using generalized partial volume joint histogram estimation. Medical Imaging, IEEE Transactions on 22(9), 1111–1119 (2003)
4. Collignon, A., Maes, F., Delaere, D., Vandermeulen, D., Suetens, P., Marchal, G.: Automated multi-modality image registration based on information theory. Information Processing in Medical Imaging pp. 263–274 (1995)
5. Haber, E., Modersitzki, J.: Intensity gradient based registration and fusion of multi-modal images. Methods of information in medicine 46(3), 292–299 (2007)
6. Hermosillo, G., Chefd'Hotel, C., Faugeras, O.: Variational methods for multimodal image matching. International Journal of Computer Vision 50(3), 329–343 (2002)
7. Koenderink, J., Van Doorn, A.: The structure of locally orderless images. International Journal of Computer Vision 31(2), 159–168 (1999)
8. Loeckx, D., Maes, F., Vandermeulen, D., Suetens, P.: Comparison between parzen window interpolation and generalised partial volume estimation for nonrigid image registration using mutual information. Biomedical Image Registration pp. 206–213 (2006)
9. Loeckx, D., Slagmolen, P., Maes, F., Vandermeulen, D., Suetens, P.: Nonrigid image registration using conditional mutual information. In: Proceedings of the 20th international conference on Information processing in medical imaging. pp. 725–737. Springer-Verlag (2007)
10. Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P.: Multimodality image registration by maximization of mutual information. Medical Imaging, IEEE Transactions on 16(2), 187–198 (1997)
11. Maes, F., Vandermeulen, D., Suetens, P.: Comparative evaluation of multiresolution optimization strategies for multimodality image registration by maximization of mutual information. Medical Image Analysis 3(4), 373–386 (1999)
12. Marcus, D., Wang, T., Parker, J., Csernansky, J., Morris, J., Buckner, R.: Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. Journal of Cognitive Neuroscience 19(9), 1498–1507 (2007)
13. Modat, M., Ridgway, G., Taylor, Z., Lehmann, M., Barnes, J., Hawkes, D., Fox, N., Ourselin, S.: Fast free-form deformation using graphics processing units. Computer Methods and Programs in Biomedicine (2009)
14. Modersitski, J.: Numerical Methods for Image Registration. Oxford University Press (2004)
15. Pennec, X., Stefanescu, R., Arsigny, V., Fillard, P., Ayache, N.: Riemannian Elasticity: A Statistical Regularization Framework for Non-linear Registration. MICCAI 3750, 943 (2005)
16. Pluim, J., Antoine Maintz, J., Viergever, M.: Interpolation artefacts in mutual information-based image registration. Computer vision and image understanding 77(2), 211–232 (2000)
17. Pluim, J., Maintz, J., Viergever, M.: Mutual-information-based registration of medical images: a survey. IEEE transactions on medical imaging 22(8), 986–1004 (2003)

18. Shannon, C.: A mathematical theory of communication. Bell System Technical Journal 27, 379–423, 623–656 (July, October 1948)
19. Studholme, C., Hill, D., Hawkes, D.: An overlap invariant entropy measure of 3D medical image alignment. Pattern Recognition 32(1), 71–86 (1999)
20. Thevenaz, P., Unser, M.: Optimization of mutual information for multiresolution image registration. Image Processing, IEEE Transactions on 9(12), 2083–2099 (2000)
21. Wells, W., Viola, P., Atsumi, H., Nakajima, S., Kikinis, R.: Multi-modal volume registration by maximization of mutual information. Medical Image Analysis 1(1), 35–51 (1996)