FACULTY OF SCIENCE
UNIVERSITY OF COPENHAGEN

Jacobians for Lebesgue Registrations for a range of Sim ilarity Measures

Jon Sporring
Sune Darkner

T echnical Report no. 04 / 2011

ISSN: 0107-8283

Jacobians for Lebesgue Registrations for a range of Similarity Measures

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

May 2, 2011

Abstract

In [Darkner and Sporring, 2011] was presented a framework based on locally orderless images and Lebesgue integration resulting in a fast algorithm for registration using normalized mutual information as dissimilarity measure. This report extends the algorithm to arbitrary complex similarity measures and supplies the full derivatives of a range of common dissimilarity measures as well as their obvious extensions.
Keywords: Similarity measure, registration, Lebesgue integration, density estimation, scale space, lo- cally orderless images.

1 Introduction

Similarity measures are the cornerstones of image registration. They define the distance between two images in a given mutual configuration. The most common measure is the sum of squared diferences, which is often the default choice, since it is fast and fairly intuitive. However, it is not the preferred choice for medical image registration, and many measures have been investigated in the literature, each often requiring their own special implementation. In this paper, we extend recent work on unifying methodologies [Hermosillo et al., 2002, Darkner and Sporring, 2011] for linear and nonlinear functions of the intensity histograms. We use Locally orderless images (LOI) [Koenderink and Van Doorn, 1999] with an extension to joint density distributions for a wide range of similarity measures as the unifying methodology, where Lebesgue integration allow us to treat derivatives in measurement, integration, and intensity space in a well-posed manner, as well as ofer a scale-space formulation of these spaces. We illustrate with p-loss, p-Huber-loss, p-Hinge-loss, p-truncated-loss, Normalized Mutual information, p-cross correlation and p-correlation ratio.

2 Image registration

Image registration is the process of transforming one image $\Gamma : \Omega \to \Gamma$, 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 $\Phi(\Gamma, R)$ is minimized. We consider difeomorphic transformation of $N\,M$ parameters, $\phi : \Omega \quad \mathbb{R}^{NM} \to \Omega$, in short $I = \Gamma \circ \phi$. The general form of Φ is,

$$\Phi = \mathrm{M}(I, R) + \Sigma(\phi), \tag{1}$$

where M is a (dis)similarity measure between the images and $\Sigma(\phi)$ is a regularization term. We use Rie- mannian Elasticity [Pennec et al., 2005] as described in [Darkner et al., 2011].

2.1 (Dis)similarity measures

Many similarity measures are on the form of,

$$\mathrm{M} = \int_{\Omega} F\left(x, I(x), R(x)\right) dx, \tag{2}$$

where we make the distinction between diferentials to be discussed later as the element wise diferentials and the hypervolume elements $dx = dx_1 \wedge \cdots \wedge dx_N$ used here for integration. The equivalent Lebesgue integral reads,

$$\mathrm{M}_{\text{linear}} = \int_{\Gamma^2} F(i, j) h_{I,R}(i, j)\, di \wedge dj, \tag{3}$$

where h is the joint histogram or co-occurrence matrix of intensity values in I and J. Such functionals are all linear in h. Examples are, $p \geq 0$:

$$F_p(i, j) = |i - j|^p, \tag{4}$$

$$F_{\text{p-hinge}}(i, j) = \begin{cases} (|i-j| - k)^p & \text{if } |i-j| > k, \\ 0 & \text{otherwise}, \end{cases} \tag{5}$$

$$F_{\text{p-Huber}}(i, j) = \begin{cases} |i-j|^p & \text{if } |i-j| < k, \\ pk^{p-1}(i-j) - (p-1)k^p & \text{otherwise}, \end{cases} \tag{6}$$

$$F_{\text{p-trunc}}(i, j) = \begin{cases} |i-j|^p & \text{if } |i-j| < k, \\ k^p & \text{otherwise}, \end{cases} \tag{7}$$

Due to linearity, any gradient of $\mathrm{M}_{\text{linear}}$ will be independent of F.

Possibly more popular similarity measures are non-linear functions of the histogram. In general these can be written as

$$\mathrm{M}_{\text{non-linear}} = \int_{\Gamma^2} F\left(h_{I,R}(i, j)\right) di \wedge dj, \tag{8}$$

where F now denotes some non-linear functional. As will be shown later, typical non-linearity has little influence on computation time. These measures include $\mathrm{M}_{\text{linear}}$ as well as mutual information (MI),

$$\mathrm{M}_{\text{MI}} = \mathrm{H}_I + \mathrm{H}_R - \mathrm{H}_{I,R}, \tag{9}$$

where H denotes the marginal and the joint entropy of the intensity distribution [Shannon, 1948],

$$\mathrm{H}_I = -\int_{\Gamma} p_I(i) \log p_I(i)\, di, \tag{10}$$

$$\mathrm{H}_R = -\int_{\Gamma} p_R(j) \log p_R(j)\, dj, \tag{11}$$

$$\mathrm{H}_{I,R} = -\int_{\Gamma^2} p_{I,R}(i, j) \log p_{I,R}(i, j)\, di \wedge dj, \tag{12}$$

such that

$$\Phi_{\text{MI}} = -p_I(i) \log p_I(i) - p_R(j) \log p_R(j) + p_{I,R}(i,j) \log p_{I,R}(i,j). \tag{13}$$

The distributions are obtained by normalizing the histograms to unity,

$$p(i) \quad p_h \quad (i$$

) ,

$\Gamma\, h(j)dj$

$h(i, j)$

2 (14)

$\Gamma^2\, h(k, l)\, dk \wedge dl$. (15)

Finally, the last class of similarity measures, we consider, are,

$$M_{\text{combine}} = F(M_1, M_2, \ldots, M_T), \tag{16}$$

where $F: \mathbb{R}^T \rightarrow \mathbb{R}$ is a smooth function. This includes the functionals above, $F_{non-linear} = M_1$, normalized mutual information (NMI), and cross correlation (CC),

$$M_{\text{NMI}} = \frac{H_I + H_R}{H_{I,R}}, \tag{17}$$

$$M_{\text{CC}} = \int_{\Gamma^2} \frac{(i-\mu_i)(j-\mu_j)}{\sigma_i \sigma_j} p_{I,R}\, di \wedge dj, \tag{18}$$

$$\mu_i = \int_{\Gamma^2} i p_{I,R}\, di \wedge dj, \tag{19}$$

$$\sigma_i = \sqrt{\int_{\Gamma^2} (i-\mu_i)^2 p_{I,R}\, di \wedge dj}, \tag{20}$$

and similarly for μ_j and σ_j. correlation ratio for image registration was proposed in [Roche et al., 1998] but really originates in analysis of variance and is based on the factorization of the variance into variance within the classes and the variance between the class averages. The factorization of the variance can be written as follows:

$$\int_{\Gamma} (x-\mu)^2 p(x)\, dx = \sum_j \int_{\Gamma_j} (x-\mu_j+\mu_j-\mu)^2 p_j(x)\, dx \tag{21}$$

$$= \sum_j \int_{\Gamma_j} (x-\mu_j)^2 p_j(x)\, dx + \sum_j \int_{\Gamma_j} (\mu_j-\mu)^2 p_j(x)\, dx + 2 \sum_j \int_{\Gamma_j} (x-\mu_j)(\mu_j-\mu) p_j(x)\, dx \tag{22}$$

$$= \sum_j \int_{\Gamma_j} (x-\mu_j)^2 p_j(x)\, dx + \sum_j \int_{\Gamma_j} (\mu_j-\mu)^2 p_j(x)\, dx, \tag{23}$$

where we have used that

$$\sum_j \int_{\Gamma_j} (x-\mu_j)(\mu_j-\mu) p_j(x)\, dx = \sum_j (\mu_j-\mu) \int_{\Gamma_j} (x-\mu_j) p_j(x)\, dx = 0, \tag{24}$$

as well as

$$\mu = \frac{1}{|\Gamma|} \int_{\Gamma} x p(x)\, dx, \tag{25}$$

$$\mu_j = \frac{1}{|\Gamma_j|} \int_{\Gamma_j} x p_j(x)\, dx. \tag{26}$$

Thus correlation ratio is defined as,

$$1 - \frac{\sum_j \int_{\Gamma_j} (x-\mu_j)^2 p_j(x)\, dx}{\int_{\Gamma} (x-\mu)^2 f(x)\, dx} = \frac{\sum_j \int_{\Gamma_j} (\mu_j-\mu)^2 p_j(x)\, dx}{\int_{\Gamma} (x-\mu)^2 f(x)\, dx}. \tag{27}$$

In the Lebesgue frame work, correlation ratio j is interpreted as a class rather than an intensity value. Thus the use of Parzen window must be performed carefully. Smoothing in j-direction i.e. across classes should, if applied, be done according to a natural ordering. correlation ratio becomes

$$1 - \frac{\sum_j \int_{\Gamma_j} (i-\mu_j)^2 h_{I,R}(i,j)\, di}{\int_{\Gamma^2} (i-\mu)^2 h_{I,R}(i,j)\, di \wedge dj} = \frac{\sum_j \int_{\Gamma_j} (\mu_j-\mu)^2 h_{I,R}(i,j)\, di}{\int_{\Gamma^2} (i-\mu)^2 h_{I,R}(i,j)\, di \wedge dj}. \tag{28}$$

In the context of analysis of variance we believe that in fact the test quantity between the variance within the class over the total variance is mor intuitive. In implementation it makes little diference, on which fraction we optimize but from interpretation point of view we maximize the significance of the separation by minimization of the test quantity. We therefor write:

$$\frac{\sum_j \int_{\Gamma_j} (x - \mu_j)^2 p_j(x)\, dx}{\sum_j \int_{\Gamma} (\mu_j - \mu)^2 f(x)\, dx}, \tag{29}$$

which follows an F-distribution, i.e. the fraction of 2 χ^2-distributions assuming that $p_j(x)$ and $f(x)$ are identical and normally distributed. in the Lebesgue framework this becomes

$$\frac{\sum_j \int_{\Gamma_j} (i - \mu_j)^2 h_{I,R}(i,j)\, di}{\sum_j \int_{\Gamma} (\mu_j - \mu)^2 h_{I,R}(i,j)\, di}. \tag{30}$$

In addition we write

$$\mu = \frac{1}{|\Gamma|} \int_{\Gamma} i\, h_{I,R}(i, j)\, di \wedge dj, \tag{31}$$

$$\mu_j = \frac{1}{|\Gamma_j|} \int_{\Gamma_j} i\, h_{I,R}(i, j)\, di. \tag{32}$$

If we disregard the origin in the analysis of variance and view correlation ratio and our alternative measure as fraction of central moments we can generalize this to

$$1 - \frac{\sum_j \int_{\Gamma_j} (|i - \mu_j|)^p h_{I,R}(i,j)\, di}{\int_{\Gamma} (|i - \mu|)^p h_{I,R}(i,j)\, di \wedge dj} = \frac{\sum_j \int_{\Gamma_j} (|\mu_j - \mu|)^p h_{I,R}(i,j)\, di}{\int_{\Gamma} (|i - \mu|)^p h_{I,R}(i,j)\, di \wedge dj}. \tag{33}$$

The proposed alternative generalizes to

$$\frac{\sum_j \int_{\Gamma_j} (|i - \mu_j|)^p h_{I,R}(i,j)\, di}{\sum_j \int_{\Gamma} (|\mu_j - \mu|)^p h_{I,R}(i,j)\, di}. \tag{34}$$

2.2 Histograms by locally orderless images (LOI)

All the similarity measures above have been expressed in terms of their histograms, and we will show that this allows for a unifying and fast framework for all the above registration algorithms. Our algorithm is based on Locally Orderless Images (LOI) [Koenderink and Van Doorn, 1999], which is a conceptual model of images in terms of 3 fundamental scales: the amount of smoothing along the spatial domain (image smoothing), along the intensity domain (histogram smoothing), and the size of the window for calculating local histograms (the partial volume). A local histogram of a possibly warped image is written as,

$$h_I(i, x, \mathbf{\Phi}, \alpha, \beta, \sigma) = P(I(x, \mathbf{\Phi}, \sigma) - i, \beta) * W(x, \alpha), \tag{35}$$

$$I(x, \mathbf{\Phi}, \sigma) = I(x, \mathbf{\Phi}) * K(x, \sigma), \tag{36}$$

where P is a Parzen window of intensity or tonal scale $\beta \in \mathbb{R}_+$ centered at intensity $i \in \Gamma$, W is an integration window of scale $\alpha \in \mathbb{R}_+$ and located at x, K is a spatial measurement kernel of scale $\sigma \in \mathbb{R}_+$, and I is the transformed image by transformation parameters $\mathbf{\Phi}$, and $*$ is the convolution operator taken w.r.t. the variable x. The histogram h_R is defined similarly independently of $\mathbf{\Phi}$ or equivalently with unit transformation. In [Koenderink and Van Doorn, 1999] is used $P(i, \beta) = e^{-i^2/(2\beta^2)}$, and $K(x, \sigma) = W(x, \sigma) = e^{-x^T x/(2\sigma^2)}/(2\pi\sigma^2)^{N/2}$ calling this structure the Locally Orderless Image. The distributions are obtained by normalizing to unity,

$$p_I(i|x, \mathbf{\Phi}, \alpha, \beta, \sigma) = \frac{h_I(i, x, \mathbf{\Phi}, \alpha, \beta, \sigma)}{\int_{\Gamma} h_I(j, x, \mathbf{\Phi}, \alpha, \beta, \sigma)\, dj}, \tag{37}$$

$$p_I(i|\mathbf{\Phi}, \alpha, \beta, \sigma) = \frac{1}{|\Omega|} \int p_I(i|x, \mathbf{\Phi}, \alpha, \beta, \sigma)\, dx, \tag{38}$$

Ω

assuming (conditional) independence and uniformity such that $p_I(i, x|\mathbf{\Phi}, \alpha, \beta, \sigma) = p_I(i|x, \mathbf{\Phi}, \alpha, \beta, \sigma)/|\Omega|$. The density p_R is defined in a similar manner. As [Hermosillo et al., 2002, Darkner and Sporring, 2011], we extend the concept to the joint distributions as follows,

$$h_{I,R}(i, j, x, \mathbf{\Phi}, \alpha, \beta, \sigma) = (P(I(x, \mathbf{\Phi}, \sigma) - i, \beta) P(J(x, \sigma) - j, \beta)) * W(x, \alpha), \tag{39}$$

$$p_{I,R}(i, j|x, \mathbf{\Phi}, \alpha, \beta, \sigma) = \frac{h_{I,R}(i, j, \mathbf{\Phi}, x, \alpha, \beta, \sigma)}{\int_{\Gamma^2} h_{I,R}(k, l, x, \alpha, \beta, \sigma)\, dk \wedge dl}, \tag{40}$$

$$p_{I,R}(i, j|\mathbf{\Phi}, \alpha, \beta, \sigma) = \frac{1}{|\Omega|} \int_\Omega p_{I,R}(i, j|\mathbf{\Phi}, x, \alpha, \beta, \sigma)\, dx, \tag{41}$$

assuming (conditional) independence and uniformity such that $p_{I,R}(i, j, x|\mathbf{\Phi}, \alpha, \beta, \sigma) = p_{I,R}(i, j|x, \mathbf{\Phi}, \alpha, \beta, \sigma)/|\Omega|$.

3 First Order Structure

In the following we introduce the similarity measures and the gradient of (1) w.r.t. the transformation ϕ. We use the notation of diferentials, $dg(x) = Dg(x)\, dx$, where D is the partial derivative operator also known as the Jacobian. Note that the use of dx without being paired with an integration symbol denotes a vector or matrix of diferentials, and not the wedge product of its elements. Further, we will only write up non-zero terms that depend on $d\phi$. The diferential of (1) is,

$$d\Phi = d\mathrm{M} + d\Sigma, \tag{42}$$

where arguments have been omitted for brevity. Ignoring the regularization term we focus on the diferential of the similarity measures. In the following we will ignore $d\Sigma$.

The diferential of combinations of measures is found to be,

$$d\mathrm{M}_{\text{combine}} = DF(\mathrm{M}_1, \mathrm{M}_2, \ldots, \mathrm{M}_T) \begin{pmatrix} d\mathrm{M}_1 \\ d\mathrm{M}_2 \\ \vdots \\ d\mathrm{M}_T \end{pmatrix} \tag{43}$$

In terms of the computational complexity, the combination only causes a multiplication factor T. For linear Lebesgue integrals (3), the diferential becomes independent of F and is found to be,

$$d\mathrm{M}_{\text{linear}} = \int_{\Gamma^2} F(i, j)\, dh_{I,R}\, di \wedge dj. \tag{44}$$

under the mild Leibnitz integration rule. For non-linear similarity measures (8) the diferential is found to be,

$$d\mathrm{M}_{\text{non-linear}} = \int_{\Gamma^2} dh_{I,R}(i, j) DF(h_{I,R}(i, j))\, di \wedge dj. \tag{45}$$

Using Leibniz integration rule, the diferentials of the histograms are,

$$dh_R(j, x) = 0 \tag{46}$$

$$dh_I(i, x, \mathbf{\Phi}) = (dP(I(x, \mathbf{\Phi}, \sigma) - i, \beta) * W(x, \alpha)), \tag{47}$$

$$dh_{I,R}(i, j, x) = dP(I(\psi, \mathbf{\Phi}, \sigma) - i, \beta) P(J(\psi, \sigma) - j, \beta) * W(x - \psi, \alpha). \tag{48}$$

where irrelevant arguments have been omitted for brevity.

While the above reveals the gradient of linear similarity measures, the all non-linear measures needs to have their Jacobean derived individually. For the normalized mutual information we find that

$$dM_{NMI} = \frac{(dH_I + dH_R)H_{I,R} - (H_I + H_R)dH_{I,R}}{H^2_{I,R}}. \tag{49}$$

The entropy, H_R, is independent of ϕ, hence $d H_R = 0$. Further,

$$d H_I = -\int_\Gamma dp_I (\log p_I + 1)\, di, \tag{50}$$

$$d H_{I,R} = -\int_{\Gamma^2} dp_{I,R}(\log p_{I,R} + 1)\, di \wedge dj. \tag{51}$$

For the estimated distributions we find that

$$dp_I(i, \mathbf{\Phi}) = \frac{1}{|\Omega|}\int_\Omega dp_I(i|x, \mathbf{\Phi})\, dx, \tag{52}$$

$$dp_I(i|x, \mathbf{\Phi}) = \frac{dh_I(i, x, \mathbf{\Phi})}{\int_\Gamma h_I(j, x, \mathbf{\Phi}) dj} - \frac{h_I(i, x, \mathbf{\Phi}) \int_\Gamma dh_I(j, x, \mathbf{\Phi}) dj}{\left(\int_\Gamma h_I(j, x, \mathbf{\Phi}) dj\right)^2}, \tag{53}$$

$$dp_{I,R}(i, j) = \frac{1}{|\Omega|}\int_\Omega dp_{I,R}(i, j|x)\, dx, \tag{54}$$

$$dp_{I,R}(i, j|x) = \frac{dh_{I,R}(i, j, x)}{\int_{\Gamma^2} h_{I,R}(k, l, x)\, dk \wedge dl} - \frac{h_{I,R}(i, j, x) \int_{\Gamma^2} dh_{I,R}(k, l, x)\, dk \wedge dl}{\left(\int_{\Gamma^2} h_{I,R}(k, l, x)\, dk \wedge dl\right)^2}. \tag{55}$$

For cross correlation we find that

$$d\mu_i = \int_{\Gamma^2} i\, dp_{I,R}\, di \wedge dj, \tag{56}$$

$$d\sigma_i = \frac{1}{2\sigma_i}\int_{\Gamma^2} -2(i - \mu_i) d\mu_i p_{I,R} + (i - \mu_i)^2 dp_{I,R}\, di \wedge dj, \tag{57}$$

$$d M_{CC} = -\int_{\Gamma^2} \frac{d\mu_i (j - \mu_j) + d\mu_j (i - \mu_i) + d\sigma_i \sigma_j + d\sigma_j \sigma_i}{(\sigma_j \sigma_i)^2}\, di \wedge dj. \tag{58}$$

For correlation ratio we derive the derivatives in parts and the do assembly due to the complicated structure of the functional. From (33) and (34) we extend the 3 terms in $p - norm$ setting such that they become linear and diferentiable. In the Lebesgue framework we can rewrite (33) by setting

$$\sum_j \frac{1}{\Gamma_j}\int_{\Gamma_j} (|i - \mu_j|)^p h_{I,R}(i, j)\, di = \sum_j \frac{1}{\Gamma_j}\int_{\Gamma_j} \left(\left|i - \int_{\Gamma_j} \frac{k}{\Gamma_j} h_{I,R}(k, j)\, dk\right|\right)^p h_{I,R}(i, j)\, di \tag{59}$$

$$= \sum_j \frac{1}{\Gamma_j}\int_{\Gamma_j} \left(\left|\int_{\Gamma_j} \left(i - \frac{k}{\Gamma_j}\right) h_{I,R}(k, j)\, dk\right|\right)^p h_{I,R}(i, j)\, di. \tag{60}$$

Since $h_{I,R}$ is positive, we can write

$$\sum_j \frac{1}{\Gamma_j}\int_{\Gamma_j} \left(\int_{-\infty}^{\frac{k}{\Gamma_j}} \left(\left(i - \frac{k}{\Gamma_j}\right) h_{I,R}(k, j)\right) dk_j - \int_{\frac{k}{\Gamma_j}}^{\infty} \left(\left(\frac{k}{\Gamma_j} - i\right) h_{I,R}(k, j)\right) dk_j\right)^p h_{I,R}(i, j)\, di. \tag{61}$$

For μ this simplifies a little

$$\frac{1}{\Gamma}\int_\Gamma (|i - \mu|)^p h_{I,R}(i, j)\, di = \frac{1}{\Gamma}\int_\Gamma \left(\left|i - \int_\Gamma \frac{k}{\Gamma} h_{I,R}(k, j)\, dk\right|\right)^p h_{I,R}(i, j)\, di = \tag{62}$$

$$\frac{1}{\Gamma} \int_\Gamma \left(\left| \int_\Gamma (i-k)\, h_{I,R}(k,j) \right| dk \right)^p h_{I,R}(i,j)\, di. \tag{63}$$

Further, since $h_{I,R}$ is positive, we can write

$$\frac{1}{\Gamma} \int_\Gamma \left| \int_{-\infty}^{k_\Gamma} (i-k)\, h_{I,R}(k,j)\, dk - \int_{k_\Gamma}^{\infty} (k-i)\, h_{I,R}(k,j)\, dk \right|^p h_{I,R}(i,j)\, di. \tag{64}$$

These are both easily diferentiable structures, which shows that the partial derivatives are well defined everywhere. For the weighted variance between groups the denominator of (29). We first rewrite μ as,

$$\mu = \frac{1}{\Gamma}\int_{\Gamma} i\, h_{I,R}(i,j)\, di \wedge dj = \frac{1}{\Gamma}\sum_{j}\int_{\Gamma_j} i\, h_{I,R}(i,j)\, di, \tag{65}$$

where j is the class label. We can now rewrite (29) as,

$$\frac{1}{\Gamma}\sum_{j}\int_{\Gamma}(|\mu_j-\mu|)^p h_{I,R}(i,j)\, di = \frac{1}{\Gamma}\sum_{j}\int_{\Gamma}\Big(\Big|\frac{1}{\Gamma_j}\int_{\Gamma_j} i h_{I,R}(i,j)\, di - \frac{1}{\Gamma}\sum_{k}\int_{\Gamma_k} i h_{I,R}(i,k)\, di\Big|\Big)^p h_{I,R}(i,j)\, di \tag{66}$$

$$= \frac{1}{\Gamma}\sum_{j}\int_{\Gamma}\Big(\Big|\Big(\frac{1}{\Gamma_j}-\frac{1}{\Gamma}\Big)\int_{\Gamma_j} i h_{I,R}(i,j)\, di - \frac{1}{\Gamma}\sum_{k\neq j}\int_{\Gamma_k} i h_{I,R}(i,k)\, di\Big|\Big)^p h_{I,R}(i,k)\, di. \tag{67}$$

These individual integrals can be decomposed as follows,

$$\int_{\Gamma_k} i h_{I,R}(i,k)\, di = \int_{0}^{\infty} i h_{I,R}(i,k)\, di - \int_{-\infty}^{0} i h_{I,R}(i,k)\, di. \tag{68}$$

Thus, we get the denominator of (29) to be,

$$\frac{1}{\Gamma}\sum_{j}\int_{\Gamma}\Big(\Big|\Big(\frac{1}{\Gamma_j}-\frac{1}{\Gamma}\Big)\Big(\int_{0}^{\infty} i h_{I,R}(i,k)\, di - \int_{-\infty}^{0} i h_{I,R}(i,k)\, di\Big) - \frac{1}{\Gamma}\sum_{k\neq j}\Big(\int_{0}^{\infty} i h_{I,R}(i,k)\, di - \int_{-\infty}^{0} i h_{I,R}(i,k)\, di\Big)\Big|\Big)^p h_{I,R}(i,k)\, di, \tag{69}$$

which is a linear structure in its internal, thus easy to derive the first order structure from. We are now able to write the first order structure of generalized correlation ratio (33) and the alternative (34). As we have decomposed the parts in the fractions (33) and (34), we start out by writing the derivatives of the three unique parts of the equation:

$$dM_1 = d\sum_j \frac{1}{\Gamma_j}\int_{\Gamma_j}\Big(\int_{-\infty}^{\frac{k}{\Gamma_j}}\Big(\Big(i-\frac{k}{\Gamma}\Big) h_{I,R}(k,j)\Big) - \int_{\frac{k}{\Gamma_j}}^{\infty}\Big(\Big(\frac{k}{\Gamma}-i\Big) h_{I,R}(k,j)\Big)\, dk\Big)^p h_{I,R}(i,j)\, di$$

$$= \sum_j \frac{1}{\Gamma_j}\int_{\Gamma_j} p\Big(\int_{-\infty}^{\frac{k}{\Gamma_j}}\Big(\Big(i-\frac{k}{\Gamma}\Big) h_{I,R}(k,j)\Big) - \int_{\frac{k}{\Gamma_j}}^{\infty}\Big(\Big(\frac{k}{\Gamma}-i\Big) h_{I,R}(k,j)\Big)\, dk\Big)^{p-1}$$

$$\Big(\int_{-\infty}^{\frac{k}{\Gamma_j}}\Big(\Big(i-\frac{k}{\Gamma}\Big) dh_{I,R}(k,j)\Big) - \int_{\frac{k}{\Gamma_j}}^{\infty}\Big(\Big(\frac{k}{\Gamma}-i\Big) dh_{I,R}(k,j)\Big)dk\Big) h_{I,R}(i,j)\, di + \sum_j$$

$$\frac{1}{\Gamma_j}\int_{\Gamma_j}\Big(\int_{-\infty}^{\frac{k}{\Gamma_j}}\Big(\Big(i-\frac{k}{\Gamma}\Big) h_{I,R}(k,j)\Big) - \int_{\frac{k}{\Gamma_j}}^{\infty}\Big(\Big(\frac{k}{\Gamma}-i\Big) h_{I,R}(k,j)\Big)\, dk\Big)^p dh_{I,R}(i,j)\, di, \tag{70}$$

$$dM_2 = d\,\frac{1}{\Gamma}\int_{\Gamma}\Big(\int_{-\infty}^{\frac{k}{\Gamma}}\Big(i-\frac{k}{\Gamma}\Big) h_{I,R}(k,j)\, dk - \int_{\frac{k}{\Gamma}}^{\infty}\Big(\frac{k}{\Gamma}-i\Big) h_{I,R}(k,j)\, dk\Big)^p h_{I,R}(i,j)\, di$$

$$= \frac{1}{\Gamma}\int_{\Gamma} p\Big(\int_{-\infty}^{\frac{k}{\Gamma}}\Big(i-\frac{k}{\Gamma}\Big) h_{I,R}(k,j)\, dk - \int_{\frac{k}{\Gamma}}^{\infty}\Big(\frac{k}{\Gamma}-i\Big) h_{I,R}(k,j)\, dk\Big)^{p-1}$$

$$\Big(\int_{-\infty}^{\frac{k}{\Gamma}}\Big(i-\frac{k}{\Gamma}\Big) dh_{I,R}(k,j)\, dk -$$

Γ

$\underline{k}$

∞ $\left(k - i\right) dh_{I,R}(k,j)\, dk\, h_{I,R}(i,j) +$ Γ

$\underline{k}\Gamma$

p

Γ

$(i - k)\, h_{I,R}(k,j)\, dk -$ Γ ∞ $(k - i)\, h_{I,R}(k,j)\, dk$ Γ $dh_{I,R}(i, j)\, di,$ (71)

$-\infty$ $\frac{k}{\Gamma}$

and

$$dM_3 = d\,1 - \Gamma_j \; \Gamma\; (\overline{\Gamma} - 1)(\overline{1}\; j\; \Gamma \int_0^\infty ih_{I,R}(i,k)\,di - \int_{-\infty}^{0} ih_{I,R}(i,k)\,di) - {}^{p}$$

$$\dagger \int_0^\infty ih_{I,R}(i,k)\,di - \int_{-\infty}^{0} ih_{I,R}(i,k)\,di \quad h_{I,R}(i,k)\,di$$

$$\_\Gamma_{k=j} \quad \_\_ \quad \_ \quad 0 \quad -\infty$$

$$=1\; \Gamma\; -j\; \Gamma\; p\,(\Gamma - 1)(1\; j\; \Gamma \int_0^\infty ih_{I,R}(i,k)\,di - \int_{-\infty}^{0} ih_{I,R}(i,k)\,di) - {}^{p-1}$$

$$1\; \Gamma_{k=j} \int_0^\infty ih_{I,R}(i,k)\,di - \int_{-\infty}^{0} ih_{I,R}(i,k)\,di$$

$$(\Gamma \underline{-} 1)(1\; j \int_0^\infty ih_{I,R}(i,k)\,di - \Gamma \int_{-\infty}^{0} ih_{I,R}(i,k)\,di) -$$

$$1\; \Gamma_{k=j} \int_0^\infty ih_{I,R}(i,k)\,di - \int_{-\infty}^{0} ih_{I,R}(i,k)\,di \quad h_{I,R}(i,k)\,di +$$

$$1\; \Gamma_j\; \Gamma\; (\Gamma - 1)(1\; j\; \Gamma \int_0^\infty ih_{I,R}(i,k)\,di - \int_{-\infty}^{0} ih_{I,R}(i,k)\,di) - {}^{p}$$

$$1 \int_0^\infty ih_{I,R}(i,k)\,di -$$

$$ih_{I,R}(i,k)\,di$$

$$\Gamma_{k=j}\; dh_{I,R}(i,k)\,di. \int_{-\infty}^{0} \tag{72}$$

We denote:

$$M_1 = \sum_j \int_{\Gamma_j} (|i - \mu_j|)^p h_{I,R}(i,j)\,di \tag{73}$$

$$M_2 = \int_\Gamma (|i - \mu|)^p h_{I,R}(i,j)\,di \wedge dj \tag{74}$$

$$M_3 = \sum_j \int_{\Gamma_j} (|\mu_j - \mu|)^p h_{I,R}(i,j)\,di \tag{75}$$

From this we write (33) as

$$dM_{CR} = -M_1 dM_2 M^2 dM_1 M_2 - {}_2 \tag{76}$$

$$ord M_{CR} = M_2 dM_3 M^2 dM_2 M_3 - {}_2 \tag{77}$$

(78)

The first order structure of the alternative version 29 becomes

$$dM_{CRalt} = -M_3 dM_1 M^2 dM_3 M_1 - {}_3 \tag{79}$$

(80)

4 Summary

In the above we have discussed Image registration for various (dis)similarity measures. These have been formulated as Lebesgue integrals and histograms measured by locally orderless images. Further the first order structure has been evaluated for a wide range of (dis)similarity measures. The derivatives demonstrates the existence of the Jacobian of linear Lebesgue measures regardless of the smoothness of the corresponding loss function.

References

[Darkner et al., 2011] Darkner, S., Hansen, M. S., Larsen, R., and F., H. M. (2011). Efcient hyperelastic regularization for registration. In *Proc. SCIA 2011 SPLNCS*.

[Darkner and Sporring, 2011] Darkner, S. and Sporring, J. (2011). Generalized partial volume: An inferior density estimator to parzen windows for normalized mutual information. In *Information Processing in Medical Imaging (IPMI 2011)*. To appear.

[Hermosillo et al., 2002] Hermosillo, G., Chefd'Hotel, C., and Faugeras, O. (2002). Variational methods for multimodal image matching. *International Journal of Computer Vision*, 50(3):329-343.

[Koenderink and Van Doorn, 1999] Koenderink, J. and Van Doorn, A. (1999). The structure of locally orderless images. *International Journal of Computer Vision*, 31(2):159-168.

[Pennec et al., 2005] Pennec, X., Stefanescu, R., Arsigny, V., Fillard, P., and Ayache, N. (2005). Riemannian Elasticity: A Statistical Regularization Framework for Non-linear Registration. *MICCAI*, 3750:943.

[Roche et al., 1998] Roche, A., Malandain, G., Pennec, X., and Ayache, N. (1998). The Correlation Ratio as a New Similarity Measure for Multimodal Image Registration. *MICCAI*, pages 1115-1124.

[Shannon, 1948] Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27:379-423, 623-656.