Simultaneous Registration of Multiple Images: Similarity Metrics and Efficient Optimization - Supplementary Material

Christian Wachinger, *Member, IEEE*, and Nassir Navab, *Member, IEEE*

1 Gradients of Similarity Measures

In the following we state the gradients of the similarity measures - mutual information, correlation ratio, and correlation coefficient. This completes the article and gives the reader all the necessary information for implementing an efficient gradient-based optimization of the multivariate cost function. A good reference, with detailed deduction of the derivatives, is the work of Hermosillo *et al.* [1].

The gradient that we show in the following was in the article denoted by

$$\mathbf{J}_{\mathrm{SM}_{i,j}}(\mathbf{x}) = \left.\frac{\partial \mathrm{SM}(I_i(\mathbf{x}), I_j(\mathbf{x} \circ \exp(\mathbf{h})); \mathbf{p})}{\partial \mathbf{h}}\right|_{\mathbf{h}=\mathbf{0}} \quad (1)$$

$$= \nabla \mathrm{SM}(I_i(\mathbf{x}), I_j(\mathbf{x}); \mathbf{p}). \quad (2)$$

We consider I_i to be the fixed and $I_j^{\downarrow}$ to be the moving image. We add the symbol $^{\downarrow}$ to indicate the moving image, with respect to which we derivate. We define the following auxiliary variables (mean, variance) with i_1 an intensity in I_i and i_2 an intensity in $I_j^{\downarrow}$

$$\mu_1 = \int i_1 p(i_1) di_1 \quad (3)$$

$$\mu_2 = \int i_2 p(i_2) di_2 \quad (4)$$

$$\mu_{2|1} = \int i_2 \frac{p(i_1, i_2)}{p(i_1)} di_2 \quad (5)$$

$$v_1 = \int i_1^2 p(i_1) di_1 - \mu_1^2 \quad (6)$$

$$v_2 = \int i_2^2 p(i_2) di_2 - \mu_2^2 \quad (7)$$

$$v_{1,2} = \int i_1 i_2 p(i_1, i_2) d(i_1, i_2) - \mu_1 \cdot \mu_2 \quad (8)$$

$p(i_1)$ the probability for intensity i_1 in I_i and $p(i_1, i_2)$ the joint probability.

We apply a kernel-based Parzen window method for the non-parametric PDF estimation [2], working with Gaussian kernels. For the estimation of the kernel window size, several methods were proposed. One common technique considers the maximization of a pseudo likelihood [3], which has the drawback of a trivial maximum at zero [4]. Instead, a leave-one-out strategy was proposed [3], [1]. The reported Parzen window kernel size that led to best results was five [4], which we adopted in our experiments.

1.1 Mutual Information

The formula for mutual information is

$$\mathrm{MI}(I_i, I_j^{\downarrow}) = \mathrm{H}(I_i) + \mathrm{H}(I_j^{\downarrow}) - \mathrm{H}(I_i, I_j^{\downarrow}) \quad (9)$$

$$= \int_{\mathbf{R}^2} p(I_i, I_j^{\downarrow}) \log \frac{p(I_i, I_j^{\downarrow})}{p(I_i) \cdot p(I_j)} \quad (10)$$

with H the entropy. The derivation is

$$\nabla \mathrm{MI}(I_i, I_j^{\downarrow}) = G_\Psi * \frac{1}{|\Omega|} \left(\frac{\frac{\partial}{\partial I_j} p(I_i, I_j^{\downarrow})}{p(I_i, I_j^{\downarrow})} - \frac{\frac{\partial}{\partial I_j} p(I_j^{\downarrow})}{p(I_j^{\downarrow})} \right) \quad (11)$$

with the Gaußian G_Ψ and the image grid $|\Omega|$.

1.2 Correlation Ratio

The formula for correlation ratio is

$$\mathrm{CR}(I_i, I_j^{\downarrow}) = 1 - \frac{\mathbb{E}(\mathrm{Var}(I_j^{\downarrow}|I_i))}{\mathrm{Var}(I_j^{\downarrow})}. \quad (12)$$

The derivation is

$$\nabla \mathrm{CR}(I_i, I_j^{\downarrow}) = G_\Psi * \frac{\mu_2 - \mu_{2|1} + \mathrm{CR}(I_i, I_j^{\downarrow}) \cdot (i_2 - \mu_2)}{\frac{1}{2} \cdot v_2 \cdot |\Omega|}. \quad (13)$$

1.3 Correlation Coefficient

The formula for correlation coefficient is

$$\mathrm{CC}(I_i, I_j^{\downarrow}) = \frac{(I_i - \mu_1)(I_j^{\downarrow} - \mu_2)}{v_1 \cdot v_2} \quad (14)$$

and its derivation

$$\nabla \mathrm{CC}(I_i, I_j^{\downarrow}) = -\frac{2}{|\Omega|} \left[\frac{v_{1,2}}{v_2} \left(\frac{i_1 - \mu_1}{v_1} \right) + \mathrm{CC}(I_i, I_j^{\downarrow}) \left(\frac{i_2 - \mu_2}{v_2} \right) \right]. \quad (15)$$

REFERENCES

[1] G. Hermosillo, C. Chefd'Hotel, and O. Faugeras, "Variational Methods for Multimodal Image Matching," *International Journal of Computer Vision*, vol. 50, no. 3, pp. 329–343, 2002.

[2] E. Parzen, "On estimation of a probability density function and mode," *The Annals of Mathematical Statistics*, vol. 33, no. 3, pp. 1065–1076, 1962.

[3] B. Turlach, "Bandwidth selection in kernel density estimation: A review," *CORE and Institut de Statistique*, vol. 19, no. 4, pp. 1–33, 1993.

[4] E. D'Agostino, F. Maes, D. Vandermeulen, and P. Suetens, "A viscous fluid model for multimodal non-rigid image registration using mutual information," *Medical image analysis*, vol. 7, no. 4, pp. 565–575, 2003.