# Similarity Metrics and Efficient Optimization for Simultaneous Registration Supplementary Material - CVPR 2009

Christian Wachinger and Nassir Navab
Computer Aided Medical Procedures (CAMP), Technische Universität München, Germany
wachinge@in.tum.de, navab@in.tum.de

## 1 Multivariate Similarity Measures

Having $n$ images $\mathcal{I} = \{I_1, \dots, I_n\}$ and corresponding transformations $\mathbf{x}_i$, we set up a maximum likelihood framework to describe the registration process mathematically. The joint density function of interest is $p(I_1, I_2, \dots, I_n)$. Due to its high dimensionality, direct calculation is prohibitive, which makes an approximation necessary. We will consider two approximations that were recently introduced. The first is an approximation by *accumulating pair-wise estimates* (APE). The second approximation is based on the *congealing* framework.

### 1.1 APE

The pair-wise approximation is derived as follows. First, we derive a pair-wise approximation with respect to image $I_n$ using the product rule and conditional independence

$$p(I_1, \dots, I_n) \overset{\text{Prod.Rule}}{=} p(I_1, \dots, I_{n-1}|I_n) \cdot p(I_n) \tag{1}$$

$$\overset{\text{Cond.Indep.}}{=} \prod_{i=1}^{n-1} p(I_i|I_n) \cdot p(I_n) \tag{2}$$

Second, we take the $n$-th power of the joint density function and do the above derivation for each of the images, leading to

$$p(I_1, \dots, I_n)^n \overset{(2)}{=} p(I_1) \cdot \prod_{i=2}^{n} p(I_i|I_1) \cdot \ldots \cdot p(I_n) \cdot \prod_{i=1}^{n-1} p(I_i|I_n) \tag{3}$$

$$= \prod_{i=1}^{n} p(I_i) \cdot \prod_{i=1}^{n} \prod_{j \neq i} p(I_j|I_i). \tag{4}$$

Third the logarithm is applied

$$\log p(I_1, \dots, I_n)^n = n \cdot \log p(I_1, \dots, I_n) \tag{5}$$

$$= \sum_{i=1}^{n} \log p(I_i) + \sum_{i=1}^{n} \sum_{j \neq i} \log p(I_j|I_i). \tag{6}$$

And after transforming the equation

$$\log p(I_1, \dots, I_n) = \frac{1}{n} \sum_{i=1}^{n} \log p(I_i) + \frac{1}{n} \sum_{i=1}^{n} \sum_{j \neq i} \log p(I_j|I_i) \tag{7}$$

$$\approx \sum_{i=1}^{n} \sum_{j \neq i} \log p(I_j|I_i). \tag{8}$$

Following the works of Viola [1] and Roche *et al.* [2], it is possible to derive SSD, NCC, CR, and MI from $\log p(I_j|I_i)$.

## 1.2 Congealing

In the congealing framework [3], independent but *not* identical distribution of the coordinate samples are assumed

$$p(I_1, \ldots, I_n) = \prod_{s_k \in \Omega} p^k(I_1(s_k), \ldots, I_n(s_k)) \tag{9}$$

and then further i.i.d. images $\mathcal{I}$

$$p(I_1, \ldots, I_n) = \prod_{s_k \in \Omega} \prod_{i=1}^{n} p^k(I_i(s_k)). \tag{10}$$

Applying the logarithm we get to $\log p(I_1, \ldots, I_n) \approx -\sum_{s_k \in \Omega} H(\mathcal{I}(s_k))$ with $H$ the sample entropy. In our extension, we do not consider the images to be independent but assume each image $I_i$ dependent on a neighborhood of images $\mathcal{N}_i$. This leads to

$$p(I_1, \ldots, I_n) = \prod_{s_k \in \Omega} \prod_{i=1}^{n} p^k(I_i(s_k) | I_{\mathcal{N}_i}(s_k)). \tag{11}$$

The size of the neighborhood depends on the structure in the image stack. If there is no further information about the images, considering a total neighborhood seems reasonable. If there is, however, a certain ordering in the stack (camera parameters,...), a smaller neighborhood may lead to better estimates.

### 1.2.1 Voxel-wise SSD

Since the approach taken in the congealing framework focuses on certain pixel or voxel locations, it is also referred to as *voxel-wise* estimation [4]. In [5], a similarity measure called voxel-wise SSD was proposed, which does a voxel-wise estimation and combines this with the assumptions of SSD. The above introduced extension nicely allows us to derive this measure.

We assume, like for SSD, Gaußian distributed intensity values and the identity as intensity mapping. Furhter, we incorporate the neighborhood information by estimating the mean $\mu_k$ for each voxel location $s_k$ with

$$\mu_k = \frac{1}{n} \sum_{l=1}^{n} I_l(s_k). \tag{12}$$

With respect to the neighborhood $\mathcal{N}_i$, the calculation of the mean should not include the image $I_i$ itself. However, this leads to a much higher computational cost because for each image and each voxel location a different mean has to be calculated. We therefore go ahead with the approximation, leading to

$$p(I_1, \ldots, I_n) = \prod_{s_k \in \Omega} \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left( -\frac{(I_i(s_k) - \mu_k)^2}{2\sigma^2} \right) \tag{13}$$

with a Gaußian variance $\sigma^2$. This leads to the formula for voxel-wise SSD

$$\log p(I_1, \ldots, I_n) \approx - \sum_{s_k \in \Omega} \sum_{i=1}^{n} (I_i(s_k) - \mu_k)^2. \tag{14}$$

Looking at the formula, we can see that voxel-wise SSD leads to the calculation of the *variance* at each location and subsequently accumulates the values [6]. The variance is one of the measures to express the *statistical dispersion* of a random variable. In contrast to entropy, measuring the structured-ness of a variable, it can only deal with mono-modal matchings.

### 1.3 Connection between APE and congealing

We show the connection between the two approximations, by starting with the extension of congealing, equation (11), and derive the formula of APE from it

$$p(I_1,\ldots,I_n) = \prod_{s_k\in\Omega}\prod_{i=1}^{n} p^k(I_i(s_k)|I_{\mathcal{N}_i}(s_k)) \tag{15}$$

$$\overset{\text{Bayes}}{=} \prod_{s_k\in\Omega}\prod_{i=1}^{n} p^k(I_{\mathcal{N}_i}(s_k)|I_i(s_k))\frac{p^k(I_i(s_k))}{p^k(I_{\mathcal{N}_i}(s_k))} \tag{16}$$

$$\overset{\text{Cond.Indep.}}{=} \prod_{s_k\in\Omega}\prod_{i=1}^{n}\left[\prod_{j\in\mathcal{N}_i} p^k(I_j(s_k)|I_i(s_k))\right]\frac{p^k(I_i(s_k))}{p^k(I_{\mathcal{N}_i}(s_k))} \tag{17}$$

$$\overset{\text{Indep.}}{=} \prod_{s_k\in\Omega}\prod_{i=1}^{n}\left[\prod_{j\in\mathcal{N}_i} p^k(I_j(s_k)|I_i(s_k))\right]\frac{p^k(I_i(s_k))}{\prod_{j\in\mathcal{N}_i} p^k(I_j(s_k))} \tag{18}$$

applying the logarithm

$$\log p(I_1,\ldots,I_n) = \sum_{s_k\in\Omega}\sum_{i=1}^{n}\sum_{j\in\mathcal{N}_i}\left(\log p^k(I_j(s_k)|I_i(s_k)) - \log p^k(I_j(s_k))\right) + \sum_{s_k\in\Omega}\sum_{i=1}^{n}\log p^k(I_i(s_k)) \tag{19}$$

and assuming further a total neighborhood

$$\log p(I_1,\ldots,I_n) = \sum_{s_k\in\Omega}\sum_{i=1}^{n}\sum_{j\neq i}\left(\log p^k(I_j(s_k)|I_i(s_k)) - \log p^k(I_j(s_k))\right) + \sum_{s_k\in\Omega}\sum_{i=1}^{n}\log p^k(I_i(s_k)) \tag{20}$$

An assumption that is different between the pair-wise and voxel-wise (congealing), per design, is that the voxel-wise coordinate samples are not identically distributed. To relate the two approaches, we set the distribution of the coordinate samples identical

$$\log p(I_1,\ldots,I_n) = \sum_{i=1}^{n}\sum_{j\neq i}(\log p(I_j|I_i) - \log p(I_j)) + \sum_{i=1}^{n}\log p(I_i) \tag{21}$$

$$= \sum_{i=1}^{n}\sum_{j\neq i}\log p(I_j|I_i) + \sum_{i=1}^{n}\log p(I_i) - \sum_{i=1}^{n}\sum_{j\neq i}\log p(I_j) \tag{22}$$

$$= \sum_{i=1}^{n}\sum_{j\neq i}\log p(I_j|I_i) + \sum_{i=1}^{n}\log p(I_i) - (n-1)\sum_{j=1}^{n}\log p(I_j) \tag{23}$$

For comparison, the derived formula for the pair-wise approach is (reciting Equation (7))

$$\log p_{\text{pw}}(I_1,\ldots,I_n) = \sum_{i=1}^{n}\sum_{j\neq i}\log p(I_j|I_i) + \sum_{i=1}^{n}\log p(I_i) \tag{24}$$

So the pair-wise and congealing approximation are, under the assumption of a total neighborhood, conditional independent images and identical distribution of coordinate samples, equal up to the term $-(n-1)\sum_{j=1}^{n}\log p(I_j)$. This term can be neglected because it does not change during the optimization process.

## 2 Gradients of Similarity Measures

In the following we state the gradients of the similarity measures - mutual information, correlation ratio, and correlation coefficient - for which we have found no place in the main article. This completes the article and gives the reader all the necessary information for implementing an efficient gradient-based optimization of the multivariate cost function. However, we also want to state that there is no contribution in the calculation of these gradients, since they are standard, and can *e.g.* be found in the work of Hermosillo *et al.* [7].

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

$$\left.\frac{\partial \mathrm{SM}(I_i, I)}{\partial I}\right|_{I=I_j^{\downarrow}} = \nabla \mathrm{SM}(I_i, I_j^{\downarrow}) \tag{25}$$

with $I_i$ the fixed and $I_j^{\downarrow}$ the moving image. We define the following auxiliary variables (mean, variance) with $i_1$ an intensity in $I_i$ and $i_2$ an intensity in $I_j^{\downarrow}$

$$\mu_1 = \int i_1 p(i_1) di_1 \tag{26}$$

$$\mu_2 = \int i_2 p(i_2) di_2 \tag{27}$$

$$\mu_{2|1} = \int i_2 \frac{p(i_1, i_2)}{p(i_1)} di_2 \tag{28}$$

$$v_1 = \int i_1^2 p(i_1) di_1 - \mu_1^2 \tag{29}$$

$$v_2 = \int i_2^2 p(i_2) di_2 - \mu_2^2 \tag{30}$$

$$v_{1,2} = \int i_1 i_2 p(i_1, i_2) d(i_1, i_2) - \mu_1 \cdot \mu_2 \tag{31}$$

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

### 2.1 Mutual Information

The formula for mutual information is

$$\mathrm{MI}(I_i, I_j^{\downarrow}) = \mathrm{H}(I_i) + \mathrm{H}(I_j^{\downarrow}) - \mathrm{H}(I_i, I_j^{\downarrow}) \tag{32}$$

$$= \int_{\mathbf{R^2}} p(I_i, I_j^{\downarrow}) \log \frac{p(I_i, I_j^{\downarrow})}{p(I_i) \cdot p(I_j)} \tag{33}$$

with H the entropy. The derivation is

$$\nabla \mathrm{MI}(I_i, I_j^{\downarrow}) = G_\Psi * \frac{1}{|\Omega|} \left( \left( \frac{\frac{\partial}{\partial I_j} p(I_i, I_j^{\downarrow})}{p(I_i, I_j^{\downarrow})} - \frac{\frac{\partial}{\partial I_j} p(I_j^{\downarrow})}{p(I_j^{\downarrow})} \right) \right) \tag{34}$$

with the Gaußian $G_\Psi$ and the image grid $|\Omega|$.

### 2.2 Correlation Ratio

The formula for correlation ratio is

$$\mathrm{CR}(I_i, I_j^{\downarrow}) = 1 - \frac{\mathbb{E}(\mathrm{Var}(I_j^{\downarrow}|I_i))}{\mathrm{Var}(I_j^{\downarrow})}. \tag{35}$$

The derivation is

$$\nabla \mathrm{CR}(I_i, I_j^{\downarrow}) = G_\Psi * \frac{\mu_2 - \mu_{2|1} + \mathrm{CR}(I_i, I_j^{\downarrow}) \cdot (i_2 - \mu_2)}{\frac{1}{2} \cdot v_2 \cdot |\Omega|}. \tag{36}$$

### 2.3 Correlation Coefficient

The formula for correlation coefficient is

$$\mathrm{CC}(\mathbf{x}) = \frac{(I_i - \mu_1)(I_j^{\downarrow} - \mu_2)}{v_1 \cdot v_2} \tag{37}$$

and its derivation

$$\nabla \mathrm{CC}(I_i, I_j^{\downarrow}) = -\frac{2}{|\Omega|}\left[\frac{v_{1,2}}{v_2}\left(\frac{i_1 - \mu_1}{v_1}\right) + \mathrm{CC}(I_i, I_j^{\downarrow})\left(\frac{i_2 - \mu_2}{v_2}\right)\right] \tag{38}$$

## References


[1] Viola, P.A.: Alignment by Maximization of Mutual Information. Ph.d. thesis, Massachusetts Institute of Technology (1995)

[2] Roche, A., Malandain, G., Ayache, N.: Unifying maximum likelihood approaches in medical image registration. Int J of Imaging Syst and Techn **11**(1) (2000) 71–80

[3] Learned-Miller, E.G.: Data driven image models through continuous joint alignment. IEEE Trans on Pattern Analysis and Machine Intelligence **28**(2) (2006) 236–250

[4] Zöllei, L., Learned-Miller, E., Grimson, E., Wells, W.: Efficient Population Registration of 3D Data. In: Computer Vision for Biomedical Image Applications, ICCV. (2005)

[5] Wachinger, C., Wein, W., Navab, N.: Three-dimensional ultrasound mosaicing. In: MICCAI, Brisbane, Australia (2007)

[6] Wachinger, C.: Three-dimensional ultrasound mosaicing. Master's thesis, Technische Universität München (2007)

[7] Hermosillo, G., Chefd'Hotel, C., Faugeras, O.: Variational Methods for Multimodal Image Matching. International Journal of Computer Vision **50**(3) (2002) 329–343