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

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

1 Derivation of SSD from the Probabilistic Framework

We mentioned in the main article that it is possible to derive several similarity measures from the log-likelihood term $\log p(I_j|I_i)$. In order to foster intuition about these derivations, we present the deduction of SSD as an example, with further information provided in [1]–[3]. For the derivation of SSD we assume a Gaussian distribution, i.i.d. coordinate samples, and the intensity mapping to be the identity. This allows us to write:

$$\log p(I_j|I_i) = \log \prod_{\mathbf{p}\in\Omega} p(I_j(\mathbf{p})|I_i(\mathbf{p})) \tag{1}$$

$$= \sum_{\mathbf{p}\in\Omega} \log \left[\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(I_j(\mathbf{p}) - I_i(\mathbf{p}))^2}{2\sigma^2}\right)\right] \tag{2}$$

$$\propto -\frac{1}{2\sigma^2} \sum_{\mathbf{p}\in\Omega} (I_j(\mathbf{p}) - I_i(\mathbf{p}))^2 \tag{3}$$

$$\propto -\mathrm{SSD}(I_j, I_i) \tag{4}$$

with variance σ^2. The maximization of the log-likelihood function corresponds to the minimization of SSD.

2 Equivalence of Voxel-Wise SSD and APE SSD

An interesting equality exists between voxel-wise SSD and accumulated pairwise estimates (APE) for SSD:

$$\sum_{s_k\in\Omega} \sum_{i=1}^{n} (I_i(s_k) - \mu_k)^2 \overset{!}{=} \frac{1}{2n} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{s_k\in\Omega} (I_i(s_k) - I_j(s_k))^2. \tag{5}$$

We show the key steps of the proof of the equality. In figure 1, we show the deduction of the left-hand side in equation (5) and in figure 2 the right-hand side. The equality of both sides shows that equation (5) holds. This equality between APE SSD and voxel-wise SSD is a reason and motivation to investigate the general relationship between APE and Congealing.

3 Connection between APE and Congealing

We show the connection between the two approximations, by starting with the Markov-congealing, see equation (10) in main article, and derive the formula of APE from it:

$$p(I_1, \dots, I_n) \tag{11}$$

$$= \prod_{s_k\in\Omega} \prod_{i=1}^{n} p^k(I_i(s_k)|I_{\mathcal{N}_i}(s_k)) \tag{12}$$

$$\overset{\text{Bayes}}{=} \prod_{s_k\in\Omega} \prod_{i=1}^{n} p^k(I_{\mathcal{N}_i}(s_k)|I_i(s_k)) \frac{p^k(I_i(s_k))}{p^k(I_{\mathcal{N}_i}(s_k))} \tag{13}$$

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

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

Applying the logarithm and assuming a maximal neighborhood leads to

$$\log p(I_1, \dots, I_n) \tag{14}$$

$$= \sum_{s_k\in\Omega} \sum_{i=1}^{n} \sum_{j\neq i} \left(\log p^k(I_j(s_k)|I_i(s_k)) - \log p^k(I_j(s_k))\right)$$

$$+ \sum_{s_k\in\Omega} \sum_{i=1}^{n} \log p^k(I_i(s_k)). \tag{15}$$

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

$$\sum_{s_k\in\Omega}\sum_{i=1}^{n}(I_i(s_k)-\mu_k)^2 = \sum_{s_k\in\Omega}\sum_{i=1}^{n}\left[I_i^2(s_k) - 2\cdot I_i(s_k)\cdot\left(\frac{1}{n}\sum_{j=1}^{n}I_j(s_k)\right)+\frac{1}{n^2}\left(\sum_{j=1}^{n}I_j(s_k)\right)^2\right] \tag{6}$$

$$= \sum_{s_k\in\Omega}\left[\sum_{i=1}^{n}I_i^2(s_k) - \frac{2}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}I_i(s_k)I_j(s_k) + \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}I_i(s_k)I_j(s_k)\right] \tag{7}$$

$$= \sum_{s_k\in\Omega}\left[\sum_{i=1}^{n}I_i^2(s_k) - \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}I_i(s_k)I_j(s_k)\right]. \tag{8}$$

Fig. 1. Deduction of left-hand side in equation (5).

$$\frac{1}{2n}\sum_{s_k\in\Omega}\sum_{i=1}^{n}\sum_{j=1}^{n}(I_i(s_k)-I_j(s_k))^2 = \frac{1}{2n}\sum_{s_k\in\Omega}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[I_i^2(s_k)+I_j^2(s_k)-2\cdot I_i(s_k)I_j(s_k)\right] \tag{9}$$

$$= \sum_{s_k\in\Omega}\left[\sum_{i=1}^{n}I_i^2(s_k) - \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}I_i(s_k)I_j(s_k)\right]. \tag{10}$$

Fig. 2. Deduction of right-hand side in equation (5).

the coordinate samples identical

$$\begin{aligned}
&\log p(I_1,\ldots,I_n) \qquad (16)\\
&= \sum_{i=1}^{n}\sum_{j\neq i}(\log p(I_j|I_i) - \log p(I_j)) + \sum_{i=1}^{n}\log p(I_i)\\
&= \sum_{i=1}^{n}\sum_{j\neq i}\log p(I_j|I_i) + \sum_{i=1}^{n}\log p(I_i) - \sum_{i=1}^{n}\sum_{j\neq i}\log p(I_j)\\
&= \sum_{i=1}^{n}\sum_{j\neq i}\log p(I_j|I_i) + \sum_{i=1}^{n}\log p(I_i) - (n-1)\sum_{j=1}^{n}\log p(I_j)
\end{aligned}$$

Comparing this result to equation (6) of the main article, we observe that both are equivalent up to the term $-(n-1)\sum_{j=1}^{n}\log p(I_j)$. Again assuming that no prior information is available, we conclude that the approximations with APE and Markov-congealing are equal, under the consideration of a maximal neighborhood, conditional independent images and an identical distribution of coordinate samples.

4 Gradient

In this section, we provide further details about the gradient calculation in section 3.3 of the main article. We begin by stating the product of the Jacobians of the transformation

$$[\mathbf{J}_w]_{\mathbf{p}}\cdot\mathbf{J}_e = \begin{pmatrix} 0 & p_z & -p_y & 1 & 0 & 0\\ -p_z & 0 & p_x & 0 & 1 & 0\\ p_y & -p_x & 0 & 0 & 0 & 1\end{pmatrix} \tag{17}$$

for a specific point $\mathbf{p}$. Considering the Jacobian matrices not with respect to one specific point, but in their general form, we get:

- $\mathbf{J}_{\mathrm{SM}_{i,j}}$ is a $1\times N$ vector containing the derivatives of the similarity metric on the diagonal
- $\mathbf{J}_{I_j}$ is an $N\times 3N$ matrix with the image gradients in all 3 directions on the diagonals
- $\mathbf{J}_w$ is a $3N\times 12$ matrix
- $\mathbf{J}_e$ is a 12×6 matrix

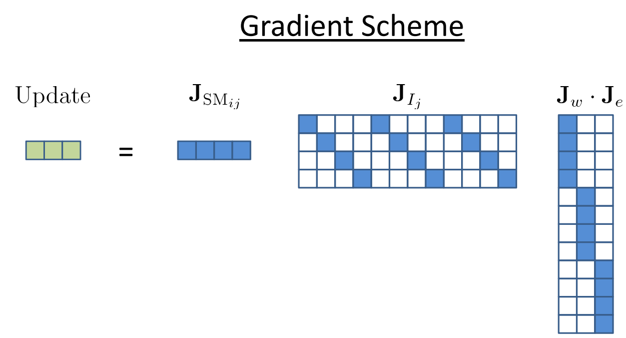
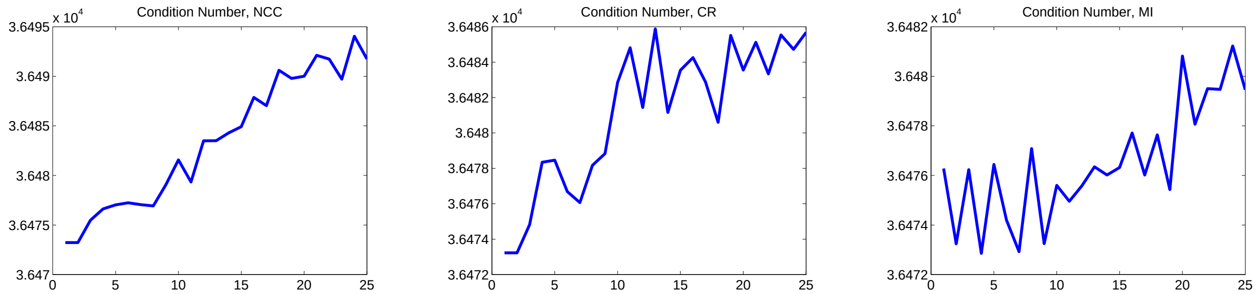


Fig. 3. Schematic illustration of the Jacobian matrices and the resulting update. Only the colored boxes are unequal zero.

with N the number of points in the grid. The product $\mathbf{J}_w\cdot\mathbf{J}_e$ is an extension of matrix in equation (60) in the main article to N points and has the following structure

$$\mathbf{J}_w\cdot\mathbf{J}_e = \begin{pmatrix} 0 & p_z^1 & -p_y^1 & 1 & 0 & 0\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ 0 & p_z^N & -p_y^N & 1 & 0 & 0\\ -p_z^1 & 0 & p_x^1 & 0 & 1 & 0\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ -p_z^N & 0 & p_x^N & 0 & 1 & 0\\ p_y^1 & -p_x^1 & 0 & 0 & 0 & 1\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ p_y^N & -p_x^N & 0 & 0 & 0 & 1\end{pmatrix}. \tag{18}$$



(a) NCC (b) CR (c) MI

Fig. 4. Plot of condition numbers for different similarity measures during optimization (x-axis: iteration number, y-axis: condition number of Hessian).

We schematically depict the scheme of the Jacobian matrices in figure 3. The illustration is corresponding to an image consisting of $N = 4$ pixels, 3 dimensions, and 3 transformation parameters (3 translations).

5 Conditioning of Hessian Approximation

We run new experiments aligning only two images. We loaded data from the RIRE dataset and applied a small transformation. We calculate the condition number of the approximated Hessian $\mathbf{J}^\top \mathbf{J}$ for each step of the optimization. The graphs for the different similarity measures are displayed in figure 4. We observe a slight increase of the condition number of the Hessian during the optimization, meaning that the sensitivity to errors increases. Following the theoretical analysis, we expected such an increase of the condition number.

References

[1] P. A. Viola, "Alignment by maximization of mutual information," Ph.D. thesis, Massachusetts Institute of Technology, 1995.

[2] A. Roche, G. Malandain, and N. Ayache, "Unifying maximum likelihood approaches in medical image registration," *International Journal of Imaging Systems and Technology: Special Issue on 3D Imaging*, vol. 11, no. 1, pp. 71–80, 2000.

[3] C. Wachinger, "Three-dimensional ultrasound mosaicing," Master's thesis, Technische Universität München, March 2007.