Groupwise Registration methods

Serdar K Balci

April 11, 2007

Abstract

We provide a comparison of groupwise registration methods. We discuss congealing, extension of pairwise to groupwise registration and use of third order entropies in groupwise registration. We give the objective functions for each method.

1 Pairwise registration

In pairwise registration, we can choose one of the subjects as a reference and register other subjects to that reference.

Let us denote the group of images as $\{u_1(x), u_2(x), \ldots, u_N(x)\}$ and associated transforms as $\{T_1(x), T_2(x), \ldots, T_N(x)\}$. Considering $u_1(x)$ as the reference image and $v_i(x) = u_i(T_i((x)))$ a pairwise approach to groupwise registration can be given as

$$\{T_2(x), \ldots, T_N(x)\} = \frac{argmax}{\{T_2(x), \ldots, T_N(x)\}} \sum_{i=2}^{N} f\left(u_1(x), u_i(T_i(x))\right) \tag{1}$$

While computing mutual information we assume that the intensities are spatially independent; however, they share the same density. The reference subject can be chosen as an intensity reference rather than an anatomical reference by allowing it to move. (This will lead to unbiased atlases?) Using MI as a similarity metric

$$S(u_1, \ldots, u_N) = \sum_{i=2}^{N} I\left(u_1(T_1(x)), u_i(T_i(x))\right) \tag{2}$$

subject to the constraint

$$\sum_{i=1}^{N} T_i(x) = 0 \tag{3}$$

The constraint allows the common coordinate frame to be at the same scale of the images. The function f can be chosen as sum of square differences, correlation coefficients or mutual information. Actually, any pairwise registration metric can be employed. Another approach to avoid choosing an anatomical reference is to register the images successively, again subject to 3.

$$S(u_1, \ldots, u_N) = \sum_{i=1}^{N} I\left(u_{i-1}(T_{i-1}(x)), u_i(T_i(x))\right) \tag{4}$$

where $u_0 = u_N$. Bhatia, uses a groupwise registration based on an extension of pairwise, however he employs a different approach. Bhatia, computes the mutual information between an intensity reference and other images. He adds all pairs of intensities in the reference and the corresponding intensity in each image to a joint histogram.

$$S(X_i r, X) = \frac{H(X) + H(X_i r)}{H(X_i r, X)} \tag{5}$$

1.1 Registering to the mean

Another method that extends pairwise to groupwise is to register each subject to the mean and update the mean after each iteration.

$$\begin{aligned} S(u_1, \ldots, u_N) &= \sum_{i=1}^{N} f\left(\mu^{k-1}(x), u_i(T_i^k(x))\right) & (6) \\ \mu^{k-1}(x) &= \frac{1}{N} \sum_{j=1}^{N} u_j(T_j^{k-1}(x)) & (7) \end{aligned}$$

where $T_i^k(x)$ refers to the i'th transform after k'th iteration and $T_i^0(x) = I(x)$

1.2 Joint pairwise registration

1.2.1 Pairwise

Instead of registering the images one to another successively we can compute the objective function value for all images jointly. The updates are computed on the sum of the pairwise objective function. The sum of the pairwise objective functions

$$\begin{aligned} S_{pair} &= \sum_{i=1}^{N} H(v_{i-1}) + H(v_i) - H(v_{i-1}, v_i) & (8) \\ &= 2\sum_{i=1}^{N} H(v_i) - \sum_{i=1}^{N} H(v_{i-1}, v_i) & (9) \end{aligned}$$

If we consider all pairs $N(N-1)/2$ pairwise registrations should be performed

$$S_{pair} = (N-1)\sum_{i=1}^{N} H(v_i) - \sum_{j>i}^{N} H(v_i, v_j) \tag{10}$$

which is same as minimizing

$$S_{pair} = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} H(v_i, v_j) \tag{11}$$

using Parzen windows to estimate the entropy

$$S_{pair} = \frac{1}{N(N-1)} \sum_{k=1}^{N} \sum_{l=k+1}^{N} \frac{-1}{N_B} \sum_{x_i \in B} \ln \left(\frac{1}{N_A} \sum_{x_j \in A} G(d_{kl}(x_i, x_j)) \right) \quad (12)$$

where

$$d_{kl}(x_i, x_j) = \frac{1}{\sigma} \sqrt{(v_k(x_i) - v_k(x_j))^2 + (v_l(x_i) - v_l(x_j))^2}$$
$$W_{kl}(x_i) = \sum_{x_j \in A} G(d_{kl}(x_i, x_j))$$

Summarizing equations (see next section for derivations)

$$S_{pair} = \frac{-1}{N_B N(N-1)} \sum_{k=1}^{N} \sum_{l=k+1}^{N} \sum_{x_i \in B} \ln \left(\frac{1}{N_A} \sum_{x_j \in A} G(d_{kl}(x_i, x_j)) \right) \quad (13)$$

$$\frac{\partial S_\mu}{\partial T_m} = \frac{1}{N N_B} \sum_{k=1, k \neq m}^{N} \sum_{x_i \in B} \frac{1}{W_{km}(x_i)} \sum_{x_j \in A} G(d_{km}(x_i, x_j)) \quad (14)$$

$$dir_{km}(x_i, x_j) \frac{\partial}{\partial T_m} (v_m(x_i) - v_m(x_j)) \quad (15)$$

$$dir_{km}(x_i, x_j) = \frac{1}{\sigma^2}(v_k(x_i) - v_k(x_j)) + \frac{1}{\sigma^2}(v_m(x_i) - v_m(x_j)) \quad (16)$$

$$d_{kl}(x_i, x_j) = \frac{1}{\sigma} \sqrt{(v_k(x_i) - v_k(x_j))^2 + (v_l(x_i) - v_l(x_j))^2} \quad (17)$$

$$W_{kl}(x_i) = \sum_{x_j \in A} G(d_{kl}(x_i, x_j)) \quad (18)$$

1.2.2 Registering to mean

Instead of picking a reference subject, we can consider the mean as the best representative of the population (or a feature) and register to the mean.

$$S_\mu = \sum_{i=1}^{N} (H(\mu) + H(v_i) - H(\mu, v_i)) \quad (19)$$
$$= N H(\mu) + \sum_{i=1}^{N} H(v_i) - \sum_{i=1}^{N} H(\mu, v_i)$$

Which is almost same as minimizing the joint entropy $H(\mu, v_i)$. Using Parzen windows to estimate the entropy

$$S_\mu = \frac{1}{N} \sum_{l=1}^{N} \frac{-1}{N_B} \sum_{x_i \in B} \ln \left(\frac{1}{N_A} \sum_{x_j \in A} G(d_l(x_i, x_j)) \right)$$

where

$$
\begin{aligned}
d_l(x_i, x_j) &= \sqrt{(\frac{\mu(x_i) - \mu(x_j)}{\sigma_\mu})^2 + (\frac{v_l(x_i) - v_l(x_j)}{\sigma})^2} \\
W_l(x_i) &= \sum_{x_j \in A} G(d_l(x_i, x_j))
\end{aligned}
$$

the derivative of the joint entropy is

$$
\frac{\partial H}{\partial T_m} = \frac{-1}{NN_B} \sum_{l=1}^{N} \sum_{x_i \in B} \frac{1}{W_l(x_i)} \sum_{x_j \in A} G(d_l(x_i, x_j)) \frac{\partial}{\partial T_m} \frac{-1}{2} (d_l(x_i, x_j))^2
$$

$$
\begin{aligned}
\frac{\partial}{\partial T_m} (d_l(x_i, x_j))^2 &= \frac{\partial}{\partial T_m} (\frac{\mu(x_i) - \mu(x_j)}{\sigma_\mu})^2 + \frac{\partial}{\partial T_m} (\frac{v_l(x_i) - v_l(x_j)}{\sigma})^2 \\
&= \frac{2}{\sigma_\mu^2} (\mu(x_i) - \mu(x_j)) \frac{\partial}{\partial T_m} (\mu(x_i) - \mu(x_j)) \\
&\quad + \frac{2}{\sigma^2} (v_l(x_i) - v_l(x_j)) \frac{\partial}{\partial T_m} (v_l(x_i) - v_l(x_j)) \\
&= \frac{2}{\sigma_\mu^2} (\mu(x_i) - \mu(x_j)) \frac{1}{N} \frac{\partial}{\partial T_m} (v_m(x_i) - v_m(x_j)) \\
&\quad + \frac{2}{\sigma^2} (v_l(x_i) - v_l(x_j)) \frac{\partial}{\partial T_m} (v_m(x_i) - v_m(x_j)) \delta(l, m) \\
&= \left(\frac{2}{N\sigma_\mu^2} (\mu(x_i) - \mu(x_j)) + \frac{2}{\sigma^2} (v_l(x_i) - v_l(x_j) \delta(l, m)) \right) \\
&\quad \frac{\partial}{\partial T_m} (v_m(x_i) - v_m(x_j))
\end{aligned}
$$

denote

$$
dir_{lm}(x_i, x_j) = \frac{1}{N\sigma_\mu^2} (\mu(x_i) - \mu(x_j)) + \frac{1}{\sigma^2} (v_l(x_i) - v_l(x_j) \delta(l, m))
$$

then the derivative is

$$
\begin{aligned}
\frac{\partial H}{\partial T_m} &= \frac{1}{NN_B} \sum_{l=1}^{N} \sum_{x_i \in B} \frac{1}{W_l(x_i)} \sum_{x_j \in A} G(d_l(x_i, x_j)) \\
&\quad dir_{lm}(x_i, x_j) \frac{\partial}{\partial T_m} (v_m(x_i) - v_m(x_j))
\end{aligned}
$$

For numerical calculations avoid $O(N^2)$ derivative evaluations

$$
\begin{aligned}
\frac{\partial H}{\partial T_m} &= \frac{1}{NN_B} \sum_{l=1}^{N} \sum_{x_i \in B} \frac{1}{W_l(x_i)} \frac{\partial v_m(x_i)}{\partial T_m} \sum_{x_j \in A} G(d_l(x_i, x_j)) dir_{lm}(x_i, x_j) \\
&\quad + \frac{1}{NN_B} \sum_{l=1}^{N} \sum_{x_i \in A} \frac{\partial v_m(x_i)}{\partial T_m} \sum_{x_j \in A} \frac{1}{W_l(x_j)} G(d_l(x_i, x_j)) dir_{lm}(x_i, x_j)
\end{aligned}
$$

Summarizing all equations

$$S_\mu = \frac{1}{N}\sum_{l=1}^{N}\frac{-1}{N_B}\sum_{x_i\in B}\ln\left(\frac{1}{N_A}\sum_{x_j\in A}G(d_l(x_i,x_j))\right) \tag{20}$$

$$\frac{\partial S_\mu}{\partial T_m} = \frac{1}{NN_B}\sum_{l=1}^{N}\sum_{x_i\in B}\frac{1}{W_l(x_i)}\sum_{x_j\in A}G(d_l(x_i,x_j)) \tag{21}$$

$$dir_{lm}(x_i,x_j)\frac{\partial}{\partial T_m}(v_m(x_i)-v_m(x_j)) \tag{22}$$

$$dir_{lm}(x_i,x_j) = \frac{1}{N\sigma_\mu^2}(\mu(x_i)-\mu(x_j))+\frac{1}{\sigma^2}(v_l(x_i)-v_l(x_j))\delta(l,m) \tag{23}$$

$$d_l(x_i,x_j) = \sqrt{(\frac{\mu(x_i)-\mu(x_j)}{\sigma_\mu})^2+(\frac{v_l(x_i)-v_l(x_j)}{\sigma})^2} \tag{24}$$

$$W_l(x_i) = \sum_{x_j\in A}G(d_l(x_i,x_j)) \tag{25}$$

2 Congealing

In congealing, we assume that in a pixel stack the intensities share the same distribution. Spatially, the distributions are assumed to be independent; however, they can have different distributions. The objective function can be given as

$$S(u_1,\ldots,u_N) = \frac{1}{M}\sum_x g\left(u_1(T_1(x)),\ldots,u_i(T_i(x))\right)) \tag{26}$$

where M is the number of samples taken from $u(x)$ and 26 is subject to the constraint 3. Using variance as an objective function 26

$$\begin{aligned}
S_\sigma &= \frac{1}{M}\sum_x\left(\frac{1}{N}\sum_{i=1}^{N}\left[u_i(T_i(x))-\mu(x)\right]^2\right)\\
&= \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{M}\sum_x\left[u_i(T_i(x))-\mu(x)\right]^2\right)\\
&= \frac{1}{N}\sum_{i=1}^{N}var(\mu(x),u_i(T_i(x)))
\end{aligned}$$

which is in the same form as 6 if we consider pairwise registration under mean square difference objective function, assuming that $\lim_{k\to\inf}\mu^k=\mu^{k-1}$.

If entropy is used as an objective function, equation 26 is

$$S_H = \frac{-1}{MN}\sum_x\sum_{i=1}^{N}\ln\frac{1}{N-1}\sum_{j=1,j\neq i}^{N}G(d(i,j))$$

where

$$d(i,j) \quad = \quad \frac{1}{\sigma} \parallel v_i(x) - v_j(x) \parallel$$

If we set $A = B$ in 62, the derivative of the objective function with respect to transform parameters is given

$$\begin{aligned}
\frac{\partial S_H}{\partial T_l} &= \frac{-1}{MN} \sum_x \sum_{i=1}^{N} \frac{1}{W(i)} \sum_{j=1, j\neq i}^{N} G(d(i,j)) \frac{\partial}{\partial T_l} d^2(i,j) \\
&= \frac{-1}{MN} \sum_x \sum_{i=1}^{N} \frac{1}{W(i)} \sum_{j=1, j\neq i}^{N} G(d(i,j)) \frac{-1}{\sigma^2} d(i,j) \frac{\partial}{\partial T_l} (v_i(x) - v_j(x)) \\
&= \frac{1}{\sigma^2 MN} \sum_x \sum_{i=1}^{N} \frac{1}{W(i)} \sum_{j=1, j\neq i}^{N} G(d(i,j)) d(i,j) \left(\frac{\partial v_i(x)}{\partial T_i} \delta(i,l) - \frac{\partial v_j(x)}{\partial T_j} \delta(j,l) \right) \\
&= \frac{1}{\sigma^2 MN} \sum_x \sum_{i=1}^{N} \frac{\partial v_i(x)}{\partial T_i} \delta(i,l) \frac{1}{W(i)} \sum_{j=1, j\neq i}^{N} G(d(i,j)) d(i,j) \\
&\quad - \frac{1}{\sigma^2 MN} \sum_x \sum_{i=1}^{N} \frac{1}{W(i)} \sum_{j=1, j\neq i}^{N} G(d(i,j)) d(i,j) \frac{\partial v_j(x)}{\partial T_j} \delta(j,l) \\
&= \frac{1}{\sigma^2 MN} \sum_x \frac{\partial v_l(x)}{\partial T_l} \frac{1}{W(l)} \sum_{j=1, j\neq l}^{N} G(d(l,j)) d(l,j) \\
&\quad - \frac{1}{\sigma^2 MN} \sum_x \sum_{i=1}^{N} \frac{1}{W(i)} G(d(i,l)) d(i,l) \frac{\partial v_l(x)}{\partial T_l} \\
&= \frac{1}{\sigma^2 MN} \sum_x \frac{\partial v_l(x)}{\partial T_l} \frac{1}{W(l)} \sum_{j=1}^{N} G(d(l,j)) d(l,j) \\
&\quad + \frac{1}{\sigma^2 MN} \sum_x \frac{\partial v_l(x)}{\partial T_l} \sum_{j=1}^{N} \frac{1}{W(j)} G(d(l,j)) d(l,j) \\
&= \frac{1}{\sigma^2 MN} \sum_x \frac{\partial v_l(x)}{\partial T_l} \sum_{j=1}^{N} \left(\frac{1}{W(l)} + \frac{1}{W(j)} \right) G(d(l,j)) d(l,j)
\end{aligned}$$

where

$$W(i) = \sum_{k=1, k\neq i}^{N} G(d(i,k))$$

summarizing the formulas

$$
\begin{aligned}
S_H &= \frac{-1}{MN}\sum_x \sum_{i=1}^{N} \ln \frac{1}{N-1} \sum_{j=1, j\neq i}^{N} G(d(i,j)) &(27)\\
\frac{\partial S_H}{\partial T_l} &= \frac{1}{\sigma^2 MN}\sum_x \frac{\partial v_l(x)}{\partial T_l} \sum_{j=1}^{N} \left(\frac{1}{W(l)} + \frac{1}{W(j)}\right) G(d(l,j))d(l,j) &(28)\\
d(i,j) &= \frac{1}{\sigma} \parallel v_i(x) - v_j(x) \parallel &(29)\\
W(i) &= \sum_{k=1, k\neq i}^{N} G(d(i,k)) &(30)
\end{aligned}
$$

3 Third order entropies

We can make use of higher order entropies

$$I(X;Y,Z) = H(X) + H(Y,Z) - H(X,Y,Z)$$

Registering each image in a pairwise fashion

$$S_{third} = \sum_{i=1}^{N} H(v_{i-2}) + H(v_{i-1}, v_i) - H(v_{i-2}, v_{i-1}, v_i) \tag{31}$$

where $u_{-1}(x) = u_{N-1}(x)$ and $u_0(x) = u_N(x)$.

3.1 Registering to mean and variance

In analogy to registering to the mean, we can make use of the information available in variance. Taking $Z = \mu(x)$ and $Y = \sigma(x)$

$$S_{\mu,\sigma} = \sum_{i=1}^{N} H(v_i) + H(\sigma, \mu) - H(v_i, \sigma, \mu) \tag{32}$$

Instead minimize $H(v_i, \sigma, \mu)$. Using Parzen windows

$$
\begin{aligned}
S_{\mu,\sigma} &= \frac{1}{N}\sum_{l=1}^{N} H(v_l, \sigma, \mu)\\
&= \frac{-1}{NN_B}\sum_{l=1}^{N}\sum_{x_i \in B} \ln \left(\frac{1}{N_A}\sum_{x_j \in A} G(d_l(x_i, x_j))\right)
\end{aligned}
$$

where $d_l(x_i, x_j)$ and $W_l(x_i)$ are

$$
\begin{aligned}
d_l(x_i, x_j) &= \sqrt{(\frac{\sigma(x_i) - \sigma(x_j)}{\sigma_\sigma})^2 + (\frac{\mu(x_i) - \mu(x_j)}{\sigma_\mu})^2 + (\frac{v_l(x_i) - v_l(x_j)}{\sigma})^2}\\
W_l(x_i) &= \sum_{x_j \in A} G(d_l(x_i, x_j))
\end{aligned}
$$

the derivative of the entropy is then

$$
\begin{aligned}
\frac{\partial H}{\partial T_m} &= \frac{-1}{NN_B}\sum_{l=1}^{N}\sum_{x_i\in B}\frac{1}{W_l(x_i)}\sum_{x_j\in A}G(d_l(x_i,x_j))\frac{\partial}{\partial T_m}\frac{-1}{2}(d_l(x_i,x_j))^2 \\
\frac{\partial}{\partial T_m}(d_l(x_i,x_j))^2 &= \frac{\partial}{\partial T_m}(\frac{\sigma(x_i)-\sigma(x_j)}{\sigma_\sigma})^2+\frac{\partial}{\partial T_m}(\frac{\mu(x_i)-\mu(x_j)}{\sigma_\mu})^2+\frac{\partial}{\partial T_m}(\frac{v_l(x_i)-v_l(x_j)}{\sigma})^2 \\
&= \frac{2}{{\sigma_\sigma}^2}(\sigma(x_i)-\sigma(x_j))\frac{\partial}{\partial T_m}(\sigma(x_i)-\sigma(x_j)) \\
&\quad +\left(\frac{2}{N\sigma_\mu^2}(\mu(x_i)-\mu(x_j))+\frac{2}{\sigma^2}(v_l(x_i)-v_l(x_j))\delta(l,m)\right) \\
&\quad \frac{\partial}{\partial T_m}(v_l(x_i)-v_l(x_j)) \\
\frac{\partial}{\partial T_m}(\sigma(x_i)-\sigma(x_j)) &= \frac{\partial}{\partial T_m}\left(\frac{1}{N}\sum_{k=1}^{N}(v_k(x_i)^2-v_k(x_j)^2)-(\mu(x_i)^2-\mu(x_j)^2)\right) \\
&= \frac{2}{N}\left((v_m(x_i)-\mu(x_i))\frac{\partial v_m(x_i)}{\partial T_l}-(v_m(x_j)-\mu(x_j))\frac{\partial v_m(x_j)}{\partial T_m}\right)
\end{aligned}
$$

opening the expression

$$
\begin{aligned}
= &\quad \frac{4}{N{\sigma_\sigma}^2}(\sigma(x_i)-\sigma(x_j))\left((v_m(x_i)-\mu(x_i))\frac{\partial v_m(x_i)}{\partial T_m}-(v_m(x_j)-\mu(x_j))\frac{\partial v_m(x_j)}{\partial T_m}\right)+ \\
&\quad \left(\frac{2}{N\sigma_\mu^2}(\mu(x_i)-\mu(x_j))+\frac{2}{\sigma^2}(v_l(x_i)-v_l(x_j))\delta(l,m)\right)\frac{\partial}{\partial T_m}(v_m(x_i)-v_m(x_j)) \\
= &\quad 2\frac{\partial v_m(x_i)}{\partial T_m} \\
&\quad \left(\frac{2}{N{\sigma_\sigma}^2}(\sigma(x_i)-\sigma(x_j))(v_m(x_i)-\mu(x_i))+\frac{2}{N\sigma_\mu^2}(\mu(x_i)-\mu(x_j))+\frac{2}{\sigma^2}(v_l(x_i)-v_l(x_j))\delta(l,m)\right) \\
&\quad -2\frac{\partial v_m(x_j)}{\partial T_m} \\
&\quad \left(\frac{2}{N{\sigma_\sigma}^2}(\sigma(x_i)-\sigma(x_j))(v_m(x_j)-\mu(x_j))+\frac{2}{N\sigma_\mu^2}(\mu(x_i)-\mu(x_j))+\frac{2}{\sigma^2}(v_l(x_i)-v_l(x_j)\delta(l,m))\right) \\
= &\quad 2dir_{lm}(x_i,x_j)\frac{\partial v_l(x_i)}{\partial T_l}+2dir_{lm}(x_j,x_i)\frac{\partial v_l(x_i)}{\partial T_l}
\end{aligned}
$$

where we define

$$
\begin{aligned}
dir_{lm}(x_i,x_j) &= \frac{2}{N{\sigma_\sigma}^2}(\sigma(x_i)-\sigma(x_j))(v_l(x_i)-\mu(x_i)) \\
&\quad +\frac{1}{N\sigma_\mu^2}(\mu(x_i)-\mu(x_j))+\frac{1}{\sigma^2}(v_l(x_i)-v_l(x_j))\delta(l,m)
\end{aligned}
$$

The summary of the equations are

$$
\begin{aligned}
S_{\mu,\sigma} &= \frac{-1}{NN_B}\sum_{l=1}^{N}\sum_{x_i\in B}\ln\left(\frac{1}{N_A}\sum_{x_j\in A}G(d_l(x_i,x_j))\right) && (33)\\
\frac{\partial S_{\mu,\sigma}}{\partial T_m} &= \frac{1}{NN_B}\sum_{l=1}^{N}\sum_{x_i\in B}\frac{1}{W_l(x_i)}\sum_{x_j\in A}G(d_l(x_i,x_j)) && (34)\\
& \left(dir_{lm}(x_i,x_j)\frac{\partial v_m(x_i)}{\partial T_m}+dir_{lm}(x_j,x_i)\frac{\partial v_m(x_i)}{\partial T_m}\right) && (35)\\
dir_{lm}(x_i,x_j) &= \frac{2}{N{\sigma_\sigma}^2}(\sigma(x_i)-\sigma(x_j))(v_m(x_i)-\mu(x_i)) && (36)\\
& +\frac{1}{N\sigma_\mu^2}(\mu(x_i)-\mu(x_j))+\frac{1}{\sigma^2}(v_l(x_i)-v_l(x_j)\delta(l,m)) && (37)\\
d_l(x_i,x_j) &= \sqrt{(\frac{\sigma(x_i)-\sigma(x_j)}{\sigma_\sigma})^2+(\frac{\mu(x_i)-\mu(x_j)}{\sigma_\mu})^2+(\frac{v_l(x_i)-v_l(x_j)}{\sigma})^2} && (38)\\
W_l(x_i) &= \sum_{x_j\in A}G(d_l(x_i,x_j)) && (39)
\end{aligned}
$$

To select $\sigma, \sigma_\mu, \sigma_\sigma$, note that if all images are normalized to a gaussian distribution then $var(\mu) = 1/N$ and $var(\sigma) = 2N$ (chi-square distributed).

3.2 Multi-Information

If we consider entropies as sets, then the sum of pairwise intersections is called Multi-Information. Ref: Shlomo Dubnov. Multi-Information is given by

$$I(x_1^n) = \sum_{i=1}^{n} H(x_i) - H(x_1, x_2, \ldots, x_n) \tag{40}$$

additional information added when one more variables are observed

$$\rho(x_1, x_2, \ldots, x_n) = I(x_1, x_2, \ldots, x_n) - I(x_1, x_2, \ldots, x_{n-1}) \tag{41}$$

Using third order entropies to register three images at the one time

$$
\begin{aligned}
S_{third} &= \sum_{i=1}^{N}\sum_{j=0}^{2}H(v_{i-j}) - \sum_{i=1}^{N}H(v_{i-2}, v_{i-1}, v_i)\\
&= 3\sum_{i=1}^{N}H(v_i) - \sum_{i=1}^{N}H(v_{i-2}, v_{i-1}, v_i) && (42)
\end{aligned}
$$

3.3 Multiple Mutual Information

Following the set interpretation of entropies, multiple mutual information defines the intersection of all sets.We note that multiple mutual information does not follow the properties of mutual information, e.g. it can be negative. Multiple mutual information is given by (ref: Tsujishita)

$$I(v_1, v_2, \dots, v_n) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{i_1 < \dots < i_k} H(v_{i_1}, \dots, v_{i_k}) \tag{43}$$

For the discussion of three term entropies see Matsuda. By Matsuda generalized mutual informations give the following contributions to the entropy

$$H(v_1, v_2, \dots, v_n) = \sum_{i=1}^{N} H(v_i) - \sum_{i<j} I_2(v_i, v_j) + \sum_{i<j<k} I_3(v_i, v_j, v_k) - \dots \tag{44}$$

First order approximation to the joint entropy corresponds to the case when we assume that all the samples are independent, e.g.

$$p(v_1, v_2, \dots, v_n) = p(v_1)p(v_2)\dots p(v_n)$$

Approximating the joint entropy using third order entropies

$$\begin{aligned} H &= \sum_{i=1}^{N} H(v_i) - \sum_{i<j} (H(v_i) + H(v_j) - H(v_i, v_j)) + \\ &\quad \sum_{i<j<k} (H(v_i) + H(v_j) + H(v_k) - H(v_i, v_j) - \\ &\quad H(v_i, v_k) - H(v_j, v_k) + H(v_i, v_j, v_k)) \end{aligned}$$

$$H = 2\sum_{i=1}^{N} H(v_i) - 2\sum_{i<j} H(v_i, v_j) + \sum_{i<j<k} H(v_i, v_j, v_k) \tag{45}$$

There are in total $N + N(N-1)/2 + N(N-1)(N-2)/6$ terms. Using third order mutual information to register three images at one time

$$S_{I_3} = \sum_{i=1}^{N} I_3(v_{i-2}, v_{i-1}, v_i) \tag{46}$$

$$S_{I_3} = \sum_{i=1}^{N} 3H(v_i) - 3H(v_i, v_{i-1}) + H(v_i, v_{i-1}, v_{i-2}) \tag{47}$$

A third order approximation to the multiple mutual information is given by

$$S_{I_N} = \sum_{i=1}^{N} H(v_i) - \sum_{i<j} H(v_i, v_j) + \sum_{i<j<k} H(v_i, v_j, v_k) \tag{48}$$

which is again $O(N^3)$

3.4 Joint Entropy

The best extension of MI into groupwise registration is to minimize the joint entropy of all data. However, to estimate the entropy we need to form N-dimensional pdf estimates. However, using parzen windows to estimate the entropy, even if the estimated pdf's are not at the best estimates, the estimated entropy and its derivatives can be a well behaved objective function which leads to good registration. Let $\vec{v}(x_i) = [v_1(x_i), \ldots, v_N(x_i)]^T$ be an N-dimensional sample through the voxel stacks. The N-dimensional entropy using Parzen window estimates

$$H(\vec{v}) = \frac{-1}{N_B} \sum_{x_i \in B} \ln \left(\frac{1}{N_A} \sum_{x_j \in A} G_\psi(\vec{v}(x_i) - \vec{v}(x_j)) \right) \tag{49}$$

Assuming $\psi = \sigma I$

$$\begin{aligned} H &= \frac{-1}{N_B} \sum_{x_i \in B} \ln \frac{1}{N_A} \left(\sum_{x_j \in A} \prod_{k=1}^{N} G_\sigma \left(v_k(x_i) - v_k(x_j) \right) \right) && (50) \\ &= const - \frac{1}{N_B} \sum_{x_i \in B} \ln \frac{1}{N_A} \sum_{x_j \in A} G(\frac{1}{\sigma} \| \vec{v}(x_i) - \vec{v}(x_j) \|) && (51) \\ & && (52) \end{aligned}$$

Which requires $O(M^2 N)$ computations, where M is the number of samples and N is the number of images. Now, consider the derivative of the entropy with respect to transform parameters.

$$\frac{\partial H}{\partial T} = \frac{-1}{N_B} \sum_{x_i \in B} \frac{\partial}{\partial T} \ln \frac{1}{N_A} \sum_{x_j \in A} G(d_{\vec{v}}(x_i, x_j))$$

The summary of the equations

$$\begin{aligned} H &= -\frac{1}{N_B} \sum_{x_i \in B} \ln \frac{1}{N_A} \sum_{x_j \in A} G(d_{\vec{v}}(x_i, x_j)) && (53) \\ \frac{\partial H}{\partial T_l} &= \frac{1}{N_B \sigma^2} \sum_{x_i \in B} \frac{1}{W(x_i)} \sum_{x_j \in A} G(d_{\vec{v}}(x_i, x_j)) && (54) \\ & \quad (v_l(x_i) - v_l(x_j)) \frac{\partial}{\partial T_l} (\vec{v}_l(x_i) - \vec{v}_l(x_j)) && (55) \\ W(x_i) &= \sum_{x_m \in A} G(d_{\vec{v}}(x_i, x_m))) && (56) \\ d_{\vec{v}}(x_i, x_j) &= \frac{1}{\sigma} \sqrt{\sum_{k=1}^{N} (v_k(x_i) - v_k(x_j))^2)} && (57) \\ & && (58) \end{aligned}$$

for numerical computations using 62

$$\frac{\partial H}{\partial T} = \frac{1}{\sigma^2 N_B}\sum_{x_i\in B}\frac{\partial\vec{v}(x_i)}{\partial T}\frac{1}{W(v_l(x_i))}\sum_{x_j\in A}G_\psi(d_{\vec{v}}(x_i,x_j))(\vec{v}(x_i)-\vec{v}_l(x_i))$$
$$-\frac{1}{\sigma^2 N_B}\sum_{x_j\in A}\frac{\partial\vec{v}(x_j)}{\partial T}\sum_{x_i\in B}\frac{1}{W(v_l(x_i))}G_\psi(d_{\vec{v}}(x_i,x_j))(\vec{v}(x_i)-\vec{v}(x_j))$$

The algorithm is quadratic in the number of samples. To decrease the number of samples used the estimate the entropy, we can use KNN to find negiboring points in the set A. We should only consider k-closest points in set A to a point in set B.

$$\frac{\partial H}{\partial T} = \frac{1}{\sigma^2 N_B}\sum_{x_i\in B}\frac{\partial\vec{v}(x_i)}{\partial T}\frac{1}{W(v_l(x_i))}\sum_{x_j\in A,\|x_j-x_i\|<R}G_\psi(d_{\vec{v}}(x_i,x_j))(\vec{v}(x_i)-\vec{v}_l(x_i))$$
$$-\frac{1}{\sigma^2 N_B}\sum_{x_j\in A,\|x_j-x_i\|<R}\frac{\partial\vec{v}(x_j)}{\partial T}\sum_{x_i\in B}\frac{1}{W(v_l(x_i))}G_\psi(d_{\vec{v}}(x_i,x_j))(\vec{v}(x_i)-\vec{v}(x_j))$$

4 Entropy estimators

4.1 Parzen Window

$$h(z) = \frac{-1}{N_B}\sum_{z_i\in B}\ln\left(\frac{1}{N_A}\sum_{z_j\in A}G_\psi(z_i-z_j)\right) \tag{59}$$

The derivative of the estimator with respect to transform parameters

$$\frac{\partial}{\partial T}h(v(T(x)) = \frac{1}{N_B}\sum_{x_i\in B}\sum_{x_j\in A}\frac{G_\psi(u_i-u_j)}{W_u(u_i)}(u_i-u_j)^T\psi^{-1}\frac{\partial}{\partial T}(u_i-u_j) \tag{60}$$

where

$$W_u(u_i) = \sum_{x_k\in A}G_\psi(u_i-u_k) \tag{61}$$

for multidimensional estimates ψ can be considered a diagonal matrix. Simplifying for numerical calculations

$$\begin{aligned}
\frac{\partial h}{\partial T} &= \frac{\psi^{-1}}{N_B}\sum_{x_i\in B}\frac{\partial}{\partial T}(u_i)\frac{1}{W(u_i)}\sum_{x_j\in A}G_\psi(u_i-u_j)(u_i-u_j)^T \qquad (62)\\
&\quad -\frac{\psi^{-1}}{N_B}\sum_{x_i\in B}\frac{1}{W(u_i)}\sum_{x_j\in A}G_\psi(u_i-u_j)(u_i-u_j)^T\frac{\partial}{\partial T}(u_j)\\
&= \frac{\psi^{-1}}{N_B}\sum_{x_i\in B}\frac{\partial}{\partial T}(u_i)\frac{1}{W(u_i)}\sum_{x_j\in A}G_\psi(u_i-u_j)(u_i-u_j)^T\\
&\quad -\frac{\psi^{-1}}{N_B}\sum_{x_j\in A}\frac{\partial}{\partial T}(u_j)\sum_{x_i\in B}\frac{1}{W(u_i)}G_\psi(u_i-u_j)(u_i-u_j)^T\\
&= \frac{\psi^{-1}}{N_B}\sum_{x_i\in B}\frac{\partial}{\partial T}(u_i)\frac{1}{W(u_i)}\sum_{x_j\in A}G_\psi(u_i-u_j)(u_i-u_j)^T\\
&\quad +\frac{\psi^{-1}}{N_B}\sum_{x_i\in A}\frac{\partial}{\partial T}(u_i)\sum_{x_j\in B}\frac{1}{W(u_i)}G_\psi(u_i-u_j)(u_i-u_j)^T
\end{aligned}$$

if $A = B$ in 62

$$\begin{aligned}
\frac{\partial h}{\partial T} &= \frac{\psi^{-1}}{N_A}\sum_{x_i\in A}\frac{\partial}{\partial T}(u_i)W(u_i)\sum_{x_j\in A}G_\psi(u_i-u_j)(u_i-u_j)^T\\
&\quad -\frac{\psi^{-1}}{N_A}\sum_{x_i\in A}\frac{\partial}{\partial T}(u_i)\sum_{x_j\in A}W(u_j)G_\psi(u_j-u_i)(u_j-u_i)^T\\
&= \frac{\psi^{-1}}{N_A}\sum_{x_i\in A}\frac{\partial}{\partial T}(u_i)W(u_i)\sum_{x_j\in A}G_\psi(u_i-u_j)(u_i-u_j)^T\\
&\quad +\frac{\psi^{-1}}{N_A}\sum_{x_i\in A}\frac{\partial}{\partial T}(u_i)\sum_{x_j\in A}W(u_j)G_\psi(u_i-u_j)(u_i-u_j)^T\\
&= \frac{\psi^{-1}}{N_A}\sum_{x_i\in A}\frac{\partial}{\partial T}(u_i)\sum_{x_j\in A}(W(u_i)+W(u_j))G_\psi(u_i-u_j)(u_i-u_j)^T
\end{aligned}$$

4.2 Order Statistics

Unbiased entropy estimator based on order statistics. Ref

$$H_{est}(z) = \frac{1}{N}\sum_{i=1}^{N-m}\log\left(\frac{N}{m}(z^{i+m}-z^i)\right) \qquad (63)$$

Lower variance estimate of entropy using order statistics

$$H_{est}(z) = \frac{1}{N-m}\sum_{i=1}^{N-m}\log(\frac{N+1}{m}(z^{i+m}-z^i)) \qquad (64)$$