Spotfinding implementation details

July 31, 2022

Dispersion Algorithm

Reducing paper equations to code form

Threshold condition for rejecting background pixels in paper (equation 2):

$$D > G\left[1 + \sigma_b \left(\frac{2}{N-1}\right)^{1/2}\right]. \tag{1}$$

Let's show how this is represented in code. With the dispersion equation

$$D = \frac{\sigma^2}{\mu}. \tag{2}$$

Expanding variance

$$\sigma^2 = \frac{1}{N-1}\sum_i (c_i - \mu)^2 \tag{3}$$

$$= \frac{1}{N-1}\left[\sum_i c_i^2 - N\mu^2\right] \tag{4}$$

$$\tag{5}$$

and with the substitutions

$$x = \sum_i c_i \tag{6}$$

$$y = \sum_i c_i^2 \tag{7}$$

$$\tag{8}$$

and given that $\mu = \frac{\sum_i c_i}{N} = \frac{x}{N}$ this can be rearranged to

$$\sigma^2 = \frac{N \cdot y - x^2}{N(N-1)}. \tag{9}$$

Dividing by the mean to calculate the dispersion then becomes:

$$D = \frac{\sigma^2}{\mu} = \frac{N}{x} \cdot \frac{N \cdot y - x^2}{N(N-1)} \tag{10}$$

$$= \frac{N \cdot y - x^2}{x(N-1)} \tag{11}$$

Algorithm as implemented in code (`https://github.com/dials/dials/blob/e7f981bdaa8ecedfbbebd30b90a56020435b1213/algorithms/image/threshold/local.h#L567-L571`) multiplies through by the $x(N-1)$ so becomes

$$N \cdot y - x^2 > G \cdot x \cdot (N-1) \left[1 + \sigma_s \cdot \sqrt{\frac{2}{N-1}}\right] \tag{12}$$

$$> G \cdot x \cdot \left[(N-1) + \sigma_b \cdot \sqrt{2(N-1)}\right] \tag{13}$$

Which gives the actual implementation in code (code splits left and right into a, c and uses m instead of N):

$$\boxed{N \cdot y - x^2 > G \cdot x \cdot \left[N - 1 + \sigma_b \cdot \sqrt{2(N-1)}\right]} \tag{14}$$

0.0.1 Signal Selection

The previous equation is for rejecting the background, by excluding kernel areas that do not have an excessive dispersion. In addition, pixels are individually selected based upon their significance compared to the pure local mean (equation 3):

$$c_i > \mu + \sigma_s \left(G\mu\right)^{1/2} \tag{15}$$

In our representation:

$$c_i > \frac{x}{N} + \sigma_s \sqrt{\frac{Gx}{N}} \tag{16}$$

Multiplying through by N and moving the reference to x to the left side:

$$Nc_i - x > \sigma_s N \sqrt{\frac{Gx}{N}} \tag{17}$$

And then the final form in the code, with b, d for left and right parts of:

$$\boxed{Nc_i - x > \sigma_s \sqrt{GxN}} \tag{18}$$

0.1 *Extended* Dispersion Algorithm