doi:10.1017/S1431927622004949
Microsc. Microanal. 28 (Suppl 1), 2022

Improvements in Speed and Hole Finding in Leginon

William J. Rice[1*] and Anchi Cheng[2]

[1.]Department of Cell Biology, NYU Langone School of Medicine, New York, NY, USA.
[2.]National Resource for Automated Molecular Microscopy, Simons Electron Microscopy Center, New York Structural Biology Center, NY, NY, USA.
* Corresponding author: william.rice@nyulangone.org

Single particle analysis of proteins and protein complexes has become a leading method for determining 3D models in the past decade [1]. For most structures, thousands of images need to be taken, from which hundreds of thousands to millions of individual particles are picked. Automated imaging has therefore become the norm for single particle analysis, and several packages are available, including Leginon [2], SerialEM [3], EPU (Thermo Fisher) and Latitude (Gatan). In most packages, the user makes an atlas of the entire grid, selects promising squares for imaging, and holes on these squares are automatically selected using a pre-defined grid or through a correlation search. "Smart" searching, using machine learning approaches, are in development but not yet in general use. Since thousands of images are taken per session, even small improvements in acquisition time or targeting accuracy are helpful.

The Leginon software package uses a correlation-based hole finder to find holes suitable for imaging from subsquare images. In this approach, the user defines the size of the holes in pixels and the software does a correlation-based search using a generic soft circle for the template. The correlation image is then thresholded to a binary image to highlight only the peaks, called "blobs", and these blobs are converted to a numpy array for further analysis. The mean and standard deviation of the pixels covered by the blob are also calculated and used to filter out targets which are too thin or too thick. For images with high contrast, such as with gold foil or with thick carbon, this works well. However, for images with lower contrast, as are often obtained from thin carbon samples, one often obtains false correlation peaks (Figure 1). Thresholding may not remove these false peaks. Leginon implements a user defined lattice spacing which can help filter out false peaks, but if there are too many, such as between every hole, this filtering algorithm can fail to identify the correct peaks.

Since the holes are expected to be round and the correlation template is also round, the true thresholded peaks should also be round. In contrast, false peaks often have a decidedly non-round shape. One way to define roundness is the following ratio:

$$R = \frac{4\,\pi\,\times\,Area}{(perimeter)^2}$$

For a perfect circle, the ratio is 1. For other shapes, it is less than 1. For example, a square is 0.785, while an ellipse with major axis equal to twice the minor axis is 0.8. In Leginon, where the blobs are represented as a binary 2D numpy array, with active pixels having the value 1 and empty pixels 0, area is calculated from the sum of the array. To determine the perimeter, we do the following calculation for each pixel making up the blob:

Count the number of adjacent pixels with the value 1 (up, down, left, and right).
Subtract this from 4

Due to the small number of pixels defining the shape, a pixelation correction is needed (Figure 2). For corner pixels, which have 2 edges, using $\sqrt{2}$ for the perimeter gives values closer to the expected result and makes the minimum value easier to determine. A "roundness" parameter has been added to the holefinder function, and in practice a minimum value of 0.8-0.9 filters out most false peaks.

A speed-up has been added to the Leginon codebase which particularly helps for imaging in super-resolution mode. Leginon already includes a speed-up which saves an 8x8 pixel fake image when movies are taken, but it calculates the statistics from the full summed image. In the case of the Gatan K3 camera, the full super resolution image is 11520 x 8184 pixels, and the statistical calculations take over 1 second to determine. To speed this up, code was added to decimate images of size 4096x4096 and larger to approximately 1k x 1k. This saves about 1 second per image in the case of K3 super-resolution images. These calculations are performed on the camera PC and so should not be dependent on the workstation used to run the Leginon host.

Leginon is also used to measure ice thickness during collection. There are two methods available, one using the energy filter and a second using aperture limited scattering (ALS) [4]. The energy filter method is more robust since it does not rely on beam brightness measured at the start of the session, but it is much slower since it requires two extra images to be taken. At NYU, we run it every 400-800 images as a check for the ALS measurement. With images taken in super-resolution mode, the method is particularly slow since these large images must be returned to the Leginon host. Code was added to select a binning for these images, down to binned by 8, to speed this up. When taking super-resolution images, this reduces the extra time required for this step from 25 s to 8.1 s. While it would still be wasteful to measure thickness this way for every image, the shortened time makes occasional measurement much more palatable.

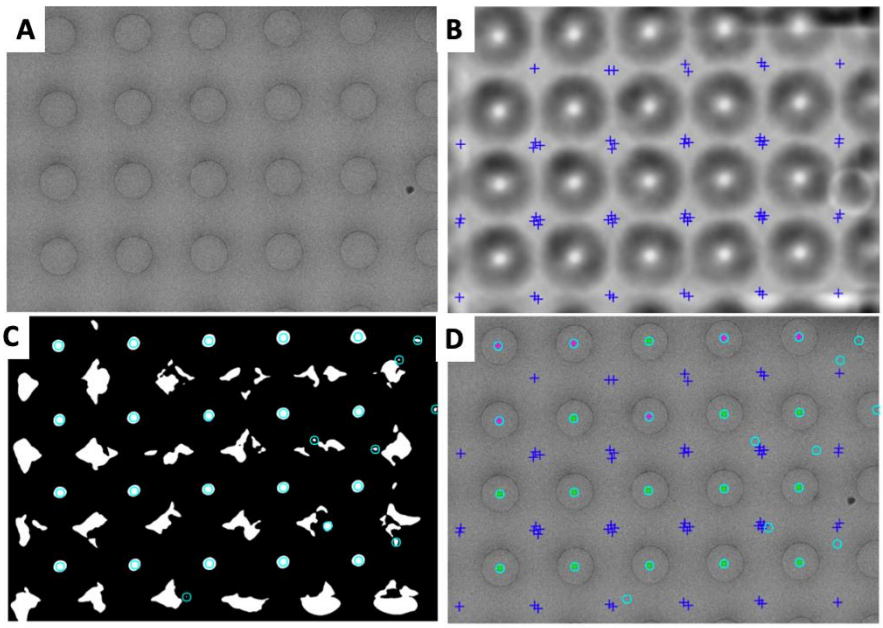

Figure 1. Correlation-based hole finding on a difficult image. A. Original subsquare image on a thin carbon grid. B. Correlation image. C. Thresholded correlation image with many false peaks. Peaks

https://doi.org/10.1017/S1431927622004949 Published online by Cambridge University Press

identified by the holefinder using the roundness criterion are marked with blue circles. D. Final targeting. Exposure targets are green crosses; focus targets are blue crosses.

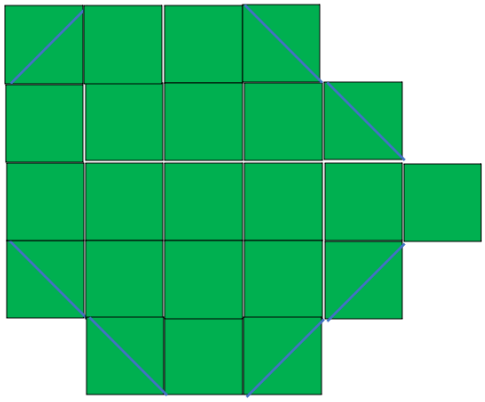

Figure 2. A small sample blob which is relatively round. The area is 23 pixels. The perimeter, as measured by counting all edges, is 22. This gives a roundness measure of 0.6. However, if 2-edged pixels are considered to contribute $\sqrt{2}$ to the perimeter (blue lines), the perimeter measures 15.3 pixels and the roundness measure increases to 0.9.

References:

[1] Y Cheng, Science **361**(6405) (2018), p. 876. doi:10.1126/science.aat4346
[2] C Suloway et al., J Struct Biol.**151** (2005), p. 41. doi: 10.1016/j.jsb.2005.03.010
[3] DN Mastronarde, J Struct Biol. **152** (2006), p.36. doi: 10.1016/j.jsb.2005.07.007
[4] WJ Rice et al., J Struct Biol. **204** (2018), p.38. doi:10.1016/j.jsb.2018.06.007

https://doi.org/10.1017/S1431927622004949 Published online by Cambridge University Press