Lesson 12: Semi-automated particle picking for BPV

In this exercise, we'll explore semi-automated particle picking using the binned BPV data which we've averaged and symmetrized in previous exercises. For simplicity, we'll use the initial, unsymmetrized average to pick particles in the same tomogram from which the initial average was created. The results are fairly typical of what you're likely to experience when applying template matching.

1) `cd $WORKSHOP_HOME/PEET_Labs/BPV_-3/PEET/templateMatching`

2) `3dmod ../../bpv_bin2.rec`
Using the ZaP or Slicer windows, re-examine this volume. Recall that it contains around 130 BPV particles at varying Z heights. We will use another modeling aid, gridInit, to prepare for template matching. Typically, you will want to choose parameters that place between 1 and 5 times as many candidate particles as there are real ones. In this case, we will end up with about 3X and we'll spread the candidate particles over 2 layers. Exit 3dmod.

3) `gridInit ../../bpv_bin2.rec 65 65 35:65:110`
Here, we've placed candidates at 65 voxel intervals in each of the 3 dimensions, leading to 450 particles. In the case of Z, I've explicitly specified start:step:end to avoid one layer being placed close to the edge of the volume. The above command puts points on Z=36 and Z=101. The default would be points at 65 and 130, and this volume is only 136 slices thick.

4) `3dmod ../../bpv_bin2.rec bpv_bin2_grid.mod`

Explore the model created by gridInit in the ZaP and Model view windows. Exit 3dmod when finished.

5) `etomo *.epe`

Notice in particular the following settings.

a) On the Setup tab, we've specified the final reference volume generated by our previous firstSearch alignment.
b) We've specified initial Uniform random rotations of the particles to minimize missing wedge artifacts, since we'll be using a limited angular search range.
c) On the Run tab, we've specified template matching by selecting No reference refinement under Optional / Advanced Features. We've also selected Strict search limit checking.
d) Since candidate particles are spaced every 65 voxels, we need an initial search distance of no less than ±32.5; I've chosen 38.
e) Finally, because we're intentionally searching far enough to allow neighboring particles to possibly converge to the same location and orientation, I've turned on duplicate removal on the Run tab. To prevent partially overlapping particles, I have chosen an unusually loose duplicate shift tolerance of 20 voxels and disabled the angular criterion by setting it to 360°.

The actual alignment takes a few hours using 20-30 cpus. As usual, we have completed the run and provided the output for you. Next we'll examine these results and decide where to set a cross-correlation threshold for which of the candidate particles to accept as genuine.

6) Press **Open averages in 3dmod** and examine the variation in the resulting averages with number of particles. The initial average with 100 particles looks reasonable. Subsequent averages aren't bad, but become increasingly noisy. This is expected, since there are only approximately 130 real particles in this tomogram. Exit Etomo and answer **Yes** to end all 3dmod programs.

7) Another way to characterize the results is to look at a histogram of the cross-correlation scores. The output motive list is in a simple .csv format readable by almost any spreadsheet, graphics, or scientific programming package. Cross-correlation scores are in column 1, so creating a histogram using Excel or LibreOffice Calc is straightforward. In this case, I've already created such a histogram for you using MATLAB. Examine it by browsing to and double-clicking on scoreHistogram.pdf. You'll see a bimodal distribution, with a smaller, upper peak of good scores above 0.35 – 0.4. Hopefully, these constitute mostly real virus particles. Close the histogram when finished.

8) `createAlignedModel *.prm`

9) `3dmod ../../bpv_bin2.rec bpv_bin2_grid_Tom1_Iter3.mod`

10) Open **Edit / Object / Type** and set **Sphere radius for points** to **19**, press **Edit / Scale bar**, check **color ramp**, and set a **minimum scale bar length** of **50** pixels. Don't close the scale bar dialog or the scale bar will disappear; you can minimize it, however, if your screen feels cluttered. In addition to placing particle model points at their aligned positions, createAlignedModel stores the cross-correlation score with each point, and can display that score as a pseudo-color. Page up and down using the ZaP or Slicer windows, and you'll see that higher cross-correlation purple and red points typically look good, while blue, green, and yellow points are often questionable. To see the numeric value of the cross-correlation score for a particular point, enter **Model** mode, left-click on the point of interest, and then press **Edit / Fine Grain**. The cross-correlation score for the selected point will be displayed at the bottom of the Fine Grain dialog under **General Value**. Try selecting a few points with different colors to see their scores. Based on this examination and the previous results visualized in the histogram, I chose to accept as genuine particles with cross-correlation scores of 0.35 or

better. Save your model. Next, we'll see how to create model, motive list, and RotAxes files containing only the selected particles.

Column 20 of motive lists is used in PEET for a class ID number. In this case, duplicate removal is enabled, so particles removed as duplicates have been assigned to class -9999, while all other particles are in class 0. We'll assign all particles not flagged as duplicates and with scores of 0.35 or larger to class 1. Since we need to refer to the existing class assignment to check for duplicates, we'll do this initially in a temporary column (**V**) within a spreadsheet program.

11) First, we'll backup the original motive list
`cp bpv_MOTL_Tom1_Iter4.csv bpv_MOTL_Tom1_Iter4.csv.orig`

12) `oocalc bpv_MOTL_Tom1_Iter4.csv`
Make sure **from row** is set to **1**, only **Comma** is checked under **Separated by**, and press **OK**.

13) Select cell **V2** and enter **=IF(AND(T2=0,A2>=0.35),1,0)**. Press **Enter**.

14) Select cell **V2** again and press **Edit / copy** or **Ctrl-C**.

15) Press **Ctrl-Shift-End** to highlight **V2:V451**, and then press **Ctrl-V** or select **Edit / Paste**). Press **Yes** when asked if it is okay to overwrite cells already containing data.

This will copy the formula from **V2** to the remaining cells in this column, adjusting row and column numbers as needed. Column **V** should now have 1's marking those particles we wish to include. Verify that this is the case.

16) **V2:V451** should still be selected. Press **Ctrl-C** to copy the entire column.

17) Right-click in cell **T2**, and select **Paste Special**.

18) In the **Paste Special** dialog, under **Selection** uncheck everything except **Numbers**. Press **OK** and then **Yes**. This will replace the contents of column **T** with only the values from column **V**. It's important to paste only values and not formulae at this step; if we pasted formulae, row numbers would get adjusted relative to their new location, and the result would no longer be correct.

19) We no longer need column **V**. Left click on the label **V** at the top of this column to select it, and then press **Delete** to erase its contents.

20) Save the file; if asked, select **Use Text CSV Format**. Exit oocalc when finished.

21) Next we need to tell PEET (specifically, those PEET programs which honor the directive we're about to use) that we only wish to process particles in class 1. This can be done in either of 2, equivalent ways:
    a. Run `gedit *.prm`, add a line at the bottom **selectClassID = 1**, save the file, and exit gedit. *NOTE*: because etomo reads the prm file on startup and writes it automatically on exit (and at other times), you must never manually edit a prm file while it is also open in etomo.
    b. Run `etomo *.epe`, under **Run** tab set **Average only members of classes** to **1**, and exit etomo.

22) `createAlignedModel *.prm`
    This generates model, motive list, and RotAxes files containing only the selected particles. In 3dmod select to **File / Reload Model** to see the newly created model on *bpv_bin2.rec* and see if you are satisfied with the selection of particles. Often, minor manual editing of automatically picked particles is desirable. In this case, there are a few reasonable looking particles which have not been included and which we might want to add manually.

Programs that honor selectClassID include averageAll, createAlignedModel, the classification, and FSC / SSNR programs. Related settings includeList and excludeList allow specifying particles by number rather than class.

After using selectClassID, includeList, or excludeList it is a good idea to delete or comment out that setting once you've finished with it. `gedit *.prm` and put a # before the **selectClassID = [1]**. This will tell PEET to ignore that line. Alternatively, you can run Etomo, set **Average only members of class** to **blank**, and exit Etomo. Confusing and seemingly mysterious results can arise if you've specified a restricted (possibly no longer even existing) subset of particles and then forgotten about it!