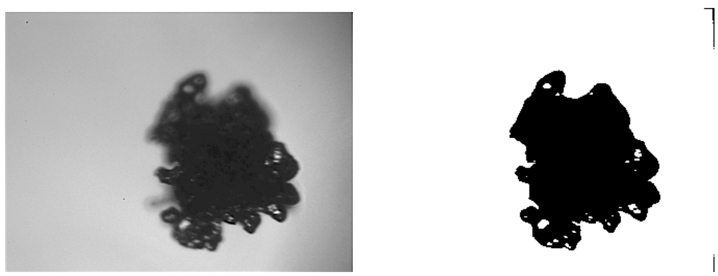
Python PHIPS Probe Habit Classification (P3HC)

This python program processes ice crystal images from the the Particle Habit Imaging and Polar Scattering probe (PHIPS), and classifies them according to their habit, following a modified version of the Holroyd (1987) classification. The structure of this program is inspired by the University of Illinois-Oklahoma Optical Array Probe Processing Software (UIOOPS).

The program requires the use of a dataset of individual raw images from the PHIPS probe must be contained in a folder, each image being in *.png* format. The function process_raw_PHIPS_images is used to process the images into simple black and white images. Regardless of initial image resolution, images are resized in processing to 256 pixels * 340 pixels, in order to remove any localized specks caused by probe graininess. Note that for the PHIPS probe, a zoom of 4x is used in typical operations, with a frame size of 2.19mm*1.65mm (Abdelmonum et al. 2016). This means that a pixel has dimensions of 6.45 μm * 6.45 μm.



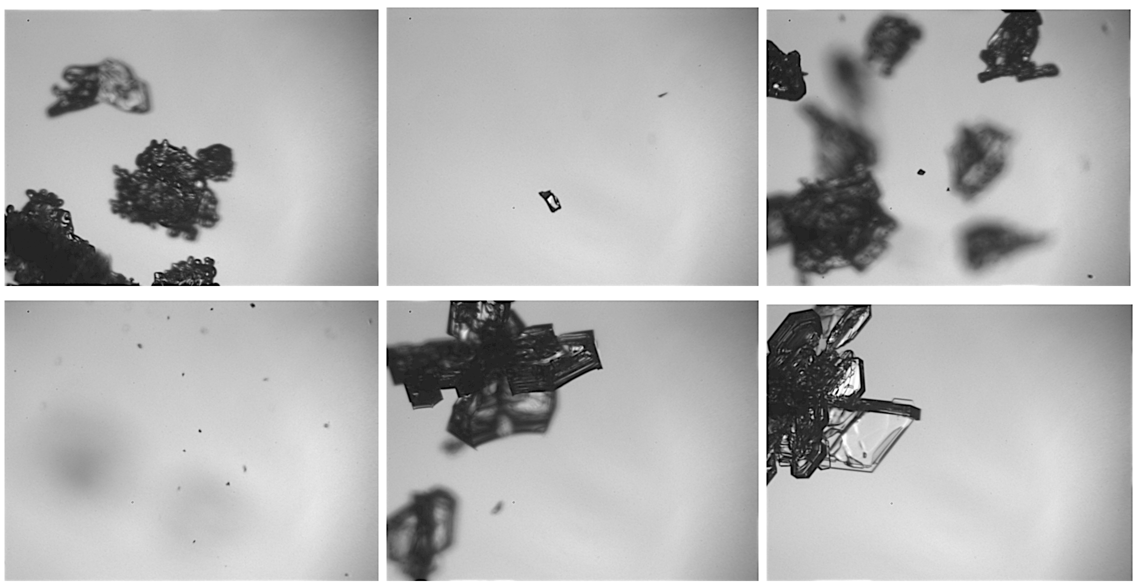
Left - Raw image; Right - Processed image.

After resizing, images are converted to black and white according to a brightness threshold, which was set at 40% for the IMPACTS campaign. Pixels with a lower brightness than the threshold are colored black, while pixels with a higher (or equal) brightness are colored white. However, the edges of the image are often darker than this threshold, despite the lack of a crystal in the region. As a result, a “buffer” of 5 pixels at the edges is ignored during further processing, since these “incorrectly” darkened pixels will often cause good images to be rejected. In other words, further processing steps act as if the image has been cropped by pixels on the edge.

Each black and white image is run through a filtering process to remove “bad” images. These include “shattered” images with many small particles, (nearly) empty images, and images where the center of the crystal is not contained within the frame. First, the horizontal and vertical dimensions of each crystal are determined by making horizontal and vertical sweeps of the image, and identifying the first and last row and column containing at least one darkened pixel. The dimensions of the crystal are approximated as the distances between the first row/column and the last row/column containing at least one darkened pixel, thus defining the smallest enclosing rectangle (given that its edges are fixed to the coordinate system). Images are rejected for being shattered artifacts if fewer than 5 percent of pixels within this smallest enclosing rectangle are darkened. Manual analysis suggests that most shattered artifacts are removed from this test, although it is also possible that very transparent crystals could mistakenly be removed as well.

Crystals that are touching the adjusted edge of the frame, after taking the five pixels closest to the edge out of consideration, can be either accepted or rejected according to a curve fit test. First, counts of the number of darkened pixels in every row and column are collected. Next, a parabolic fit is applied to the counts in each dimension. If a local maximum of darkened pixel counts is found within the frame boundaries in both dimensions, then the image is accepted; otherwise, it is rejected.

Some examples of rejected images are shown below.



For the IMPACTS campaign, the time and date of each image is contained within the filename, as well as the record number. The function leaves the time and date formatting as-is when outputting processed images into an output directory (created by the function if not yet present), but renumbers the files so that no numbers are skipped due to the removal of bad images. If the raw image files being processed have differently formatted names, the code will need to be altered accordingly.

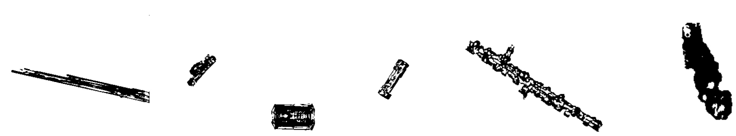
The processed images are analyzed and classified by habit using the function analyze_images. The function calculates many morphological properties of the crystal, which are defined in Table 1.

Table 1: Morphological Properties and Descriptions

Property	Description
Maximum Dimension (d)	Length of diagonal (in pixels) of smallest enclosing rectangle that is oriented in alignment with the image's coordinate system

	(used instead of smallest enclosing circle for faster processing time)
Length (x)	Number of pixels comprising the side of the smallest enclosing rectangle parallel to the x-axis (dimension with length 340 in processed images)
Width (y)	Number of pixels comprising the side of the smallest enclosing rectangle parallel to the y-axis (dimension with length 256 in processed images)
Area (a)	Number of darkened pixels in the image
Perimeter (p)	Sum of changes between dark and light pixels across each row and column
Coefficient of determination (r^2)	The coefficient of determination for the locations of all darkened pixels
Fine detail ratio (F = pd/a)	A measure of crystal "complexity"; higher values suggest a more complex and detailed crystal
Opacity coefficient (O)	The ratio of pixels darkened by any amount to pixels darkened to at least 40% brightness when reducing the image to ¼ resolution

A decision tree is used to classify habit based on these parameters, with an overall classification accuracy based on approximately 100 images of each class (except for "tiny" crystals) of 74%. These habits are described in the below table:

Habit	Description	Idealized examples (scale varies)
Tiny	Very small particles; too small to accurately assess habit	
Column	Linear crystals with one dimension greatly exceeding the other	

Aggregate	Large, complicated conglomerates of multiple crystals	
Graupel	Heavily rimed, compact, opaque pellets	
Sphere	Liquid droplets or ice crystals that are almost spherical, as evidenced by the presence of a smooth outer perimeter	
Planar	Translucent thin plates, possibly hexagonal, and simple with relatively few edges	
Dendrite	Translucent thin plates, but very complex with large numbers of edges and often quite large	
Irregular	Crystals that do not match any of the above descriptions, including rosettes	

The decisions used to classify habit are as follows:

```
if a < 100:
   Tiny
```

```
if  (r_2 >= 0.6) or (d < 150 and (x >= 2*y or y >= 2*x)) or (x >= 4*y or y >= 4*x):
   Column
if O >= 0.75 and d >= 100 and d < 250 and F <= 40:
   if F <= 15:
      Sphere
   else:
      Graupel
if d >= 250:
   if F <= 40:
      Graupel
   elif F > 150:
      Dendrite
   else:
      Aggregate
if F <= 20:
   Sphere
if F >= 200:
   Dendrite
else:
   if O < 0.35 and d >= 40:
      Planar
   else:
      Irregular
```

The values of each morphological property, and the habit class, are saved in netCDF format. All distances in pixels are converted to millimeters in the output file. Should the PHIPS probe be set to a different zoom level, adjustments in the algorithm will be necessary.

Both scripts, in addition to a script used to automate the processing, can be found using the following link: https://github.com/Raindrop57/P3HC

References

Abdelmonem, A., Järvinen, E., Duft, D., Hirst, E., Vogt, S., Leisner, T., and Schnaiter, M. (2016). PHIPS–HALO: the airborne Particle Habit Imaging and Polar Scattering probe – Part 1: Design and operation, Atmos. Meas. Tech., 9, 3131–3144, https://doi.org/10.5194/amt-9-3131-2016.

Holroyd III, E. W. (1987). Some techniques and uses of 2D-C habit classification software for snow particles. Journal of Atmospheric and Oceanic Technology, 4(3), 498-511.

UIOOPS: https://zenodo.org/record/3976291#.YtbzbS-B1D8

IMPACTS campaign website: https://espo.nasa.gov/impacts/content/IMPACTS