Machine Learning applied to Planetary Sciences

PTYS 595B/495B

Leon Palafox

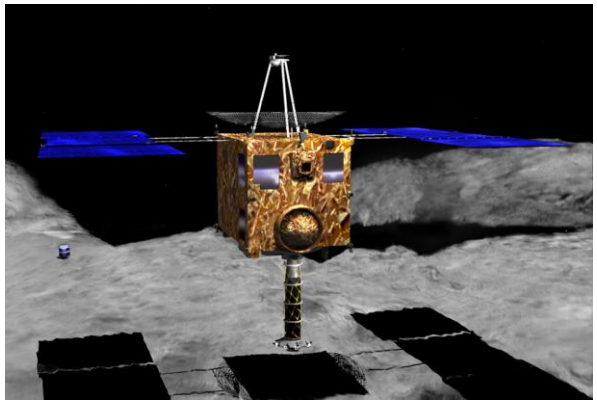
https://leonpalafox.github.io/MLClass/

Application to Boulder Detection

Fuchs, Thomas J., et al. "Enhanced flyby science with onboard computer vision: Tracking and surface feature detection at small bodies." *Earth and Space Science* 2.10 (2015): 417-434.

Problem Statement

- Hayabusa was a spacecraft that returned a small sample from an asteroid (Itokawa)
- It’s data is a reliable dataset we have for flybys in small bodies.
- We’ll be analyzing its images and dataset.

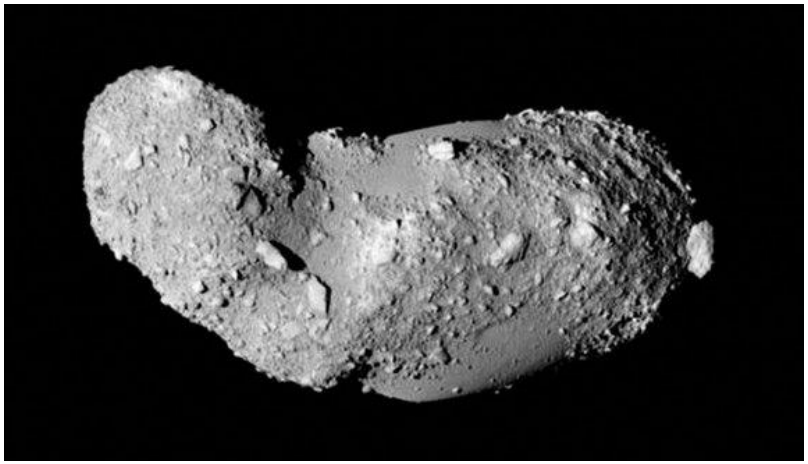


Problem Statement

- The problem with flyby boulder detection is different from traditional remote sensing:
  - You need to be able to run fast classifiers.
  - Images can be preprocessed for you, or they might not, unlike controlled remote sensing images, these are glorified photos.
  - Resolutions are not as good AND, they vary depending on the positioning of the spacecraft.

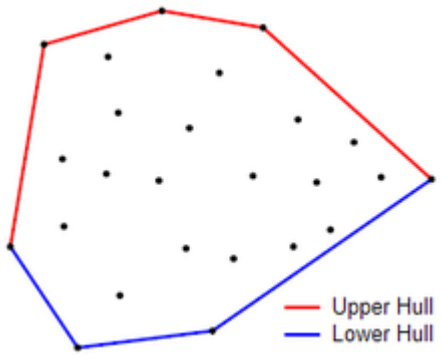
Things we don’t have to worry about

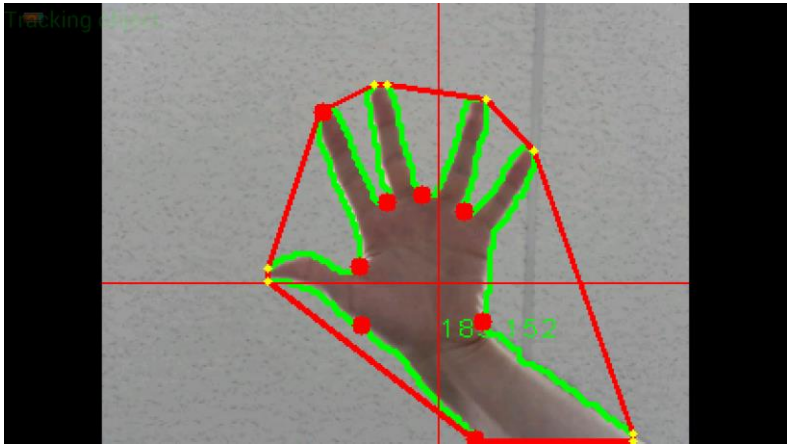
- Returning a relative set of coordinates.
- Rotation of the celestial body.
- Reattaching the mosaic dataset.
- Let’s assume tracking is solved.



First Problem with dataset

- Automated detection of the body .
  - Convex Hull
    - Family of Computer Vision algorithms to enclose a set of points.
    - It's a solved problem in CS, and both Python and Matlab have it implemented.





Second Problem of the Dataset

- There is no labeled data!
  - If there were, we wouldn't have a problem right?
- We need to use the next best thing.
- Images from previous flybys (Cassini, Stardust NeXT, etc)
  - Using these images, we need to ask experts to label them.
  - Now we pray the labels are consistent among celestial bodies.

Seriously

No Free Lunch Theorem!

- Assuming that whatever classifier we get from a different:
  - Celestial body
  - Mission
  - Instrument
  - Location in the Solar System
- Is very wishful thinking.
- Is like assuming a classifier trained on houses in Tokyo is going to work with houses in the US

It may work, it may not

Preprocessing

- If you are not using Convolutional Neural Networks.
- Is necessary to mitigate albedo changes.
  - Median filters:
    - Solved problem

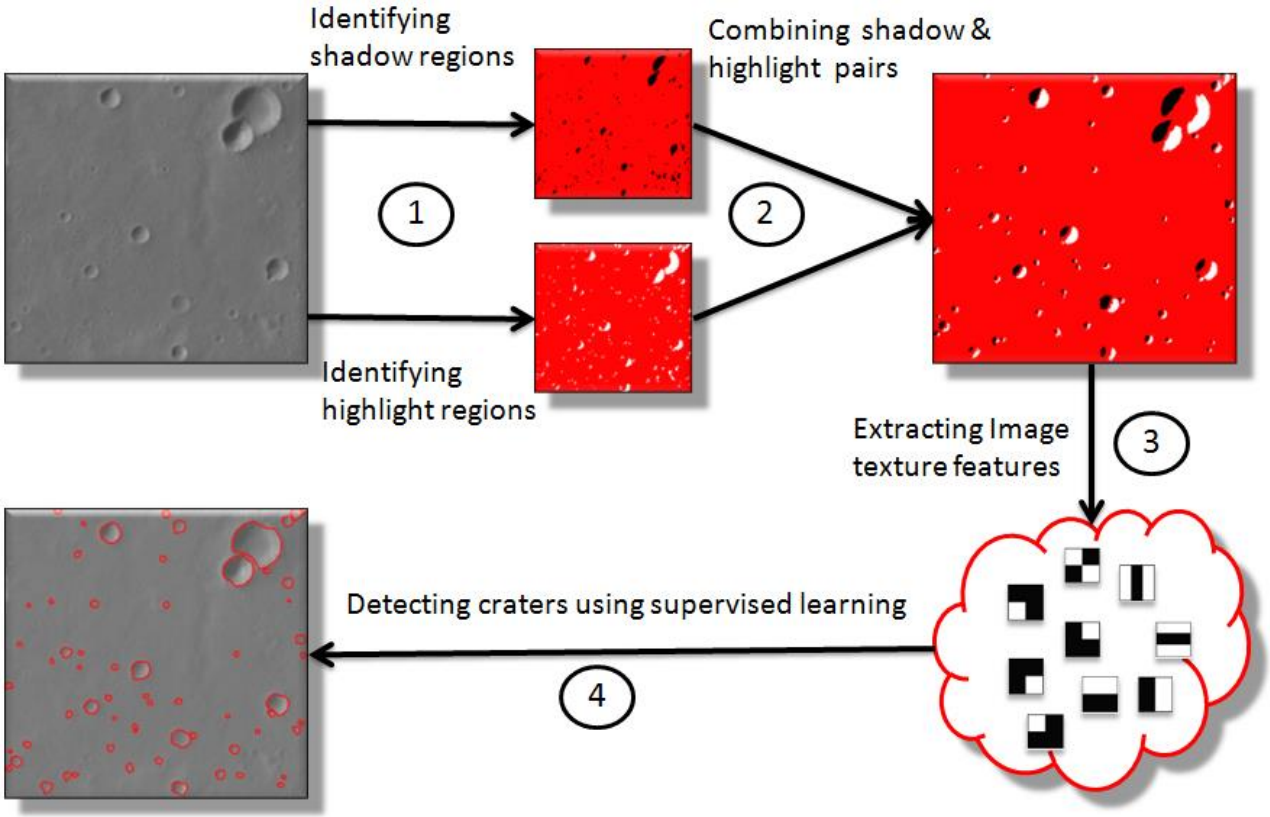
Candidate detection

- I told you this has to be fast.
- In the last class we did pixel-wise classification.
- Here we don’t care about that.
- We detect “candidate” areas that have a likelihood of having a boulder.
  - We can use this approach in VRC detection.

Candidate detection

- This is NOT a solved problem.
  - There are many papers (1340) that only deal with new methods of candidate detection.
- It can be a rough simple classification:
  - Threshold based.
  - Shape based.
  - Location based (Extremely relevant for applications)

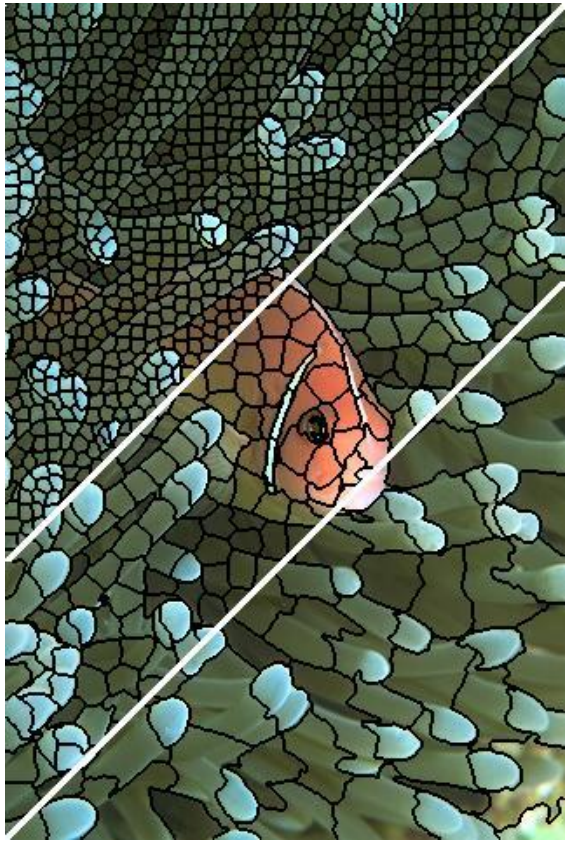
An example in crater detection



Wei Ding, Tomasz F. Stepinski, Yang Mu, et al., 2011: Subkilometer crater discovery with boosting and transfer learning, ACM TIST

Candidate Selection

- Superpixel segmentation
  - Mostly Computer Vision people.
  - Haven’t seen it used in Remote sensing much.
- Is not as fast as standard filtering techniques.



Classifier

- Papers are incredibly fuzzy on which features they use:
  - *“Finally, an attribute vector is constructed containing the raw pixel intensities, the image statistics, and the local and global histograms.”*
  - Which image statistics?
- Code is important!

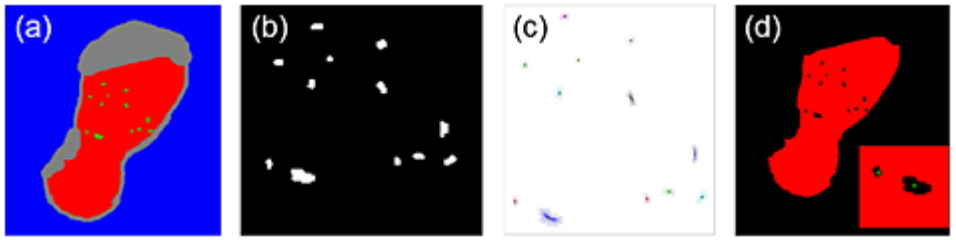
Classifier

- We need to select the window size:
  - 11x11 was the choice for this particular work.
  - They don't mention why 11? Why not 10?
    - I've been dinged for it

Classifier

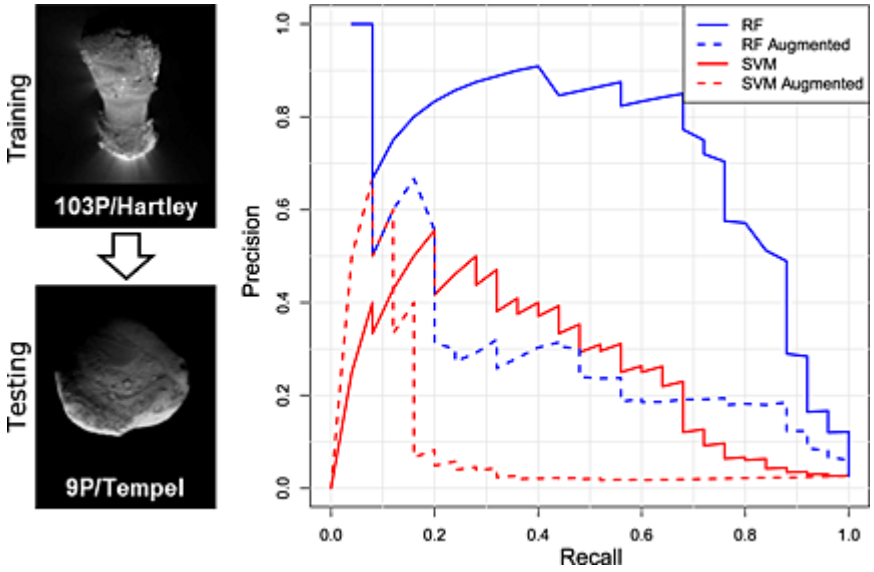
- Here they use Random Forest
  - Is a nice classifier, that can also use GPUs
  - They also compared with SVMs
    - Like in our work, SVMs is the classifier you have to compare against.
- I might have used SVMs from the beginning.

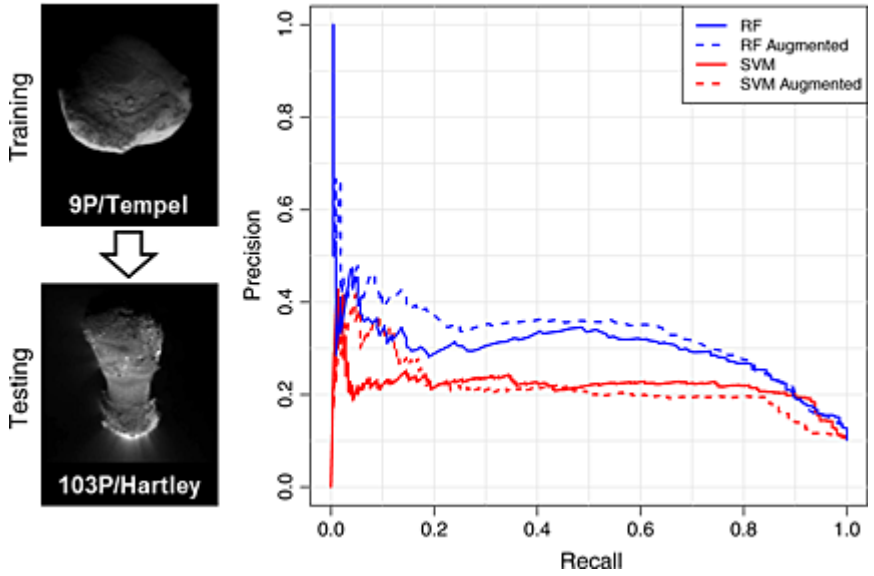
Results



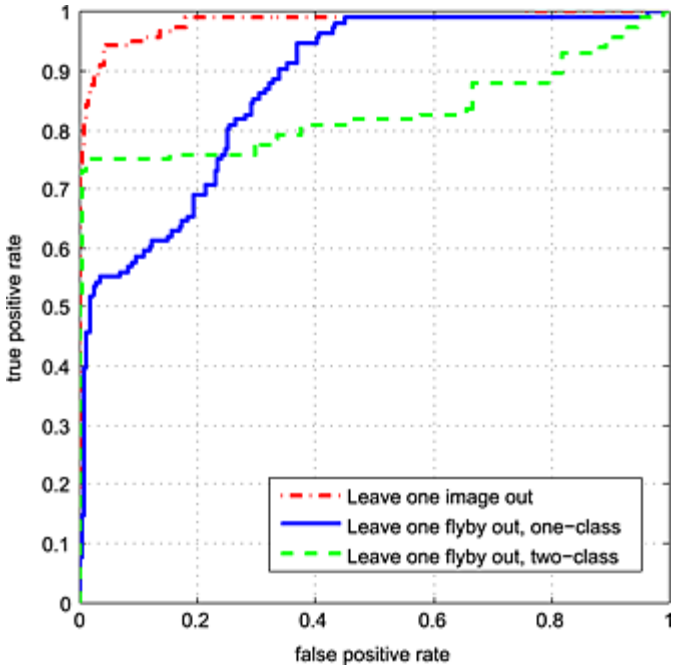
(a) Ground truth labeling of a frame from Hartley 2 in the small body data set. (b) The pixelwise labels of the surface feature layer are used for a (c) mean shift clustering to determine the centers of the blobs+. (d) The final training image for the decision forest consists of single pixel locations for the surface features (green, cf. inlay) and the surface of the small body (red).

Results


Training
103P/Hartley
Testing
9P/Tempel
Precision
Recall
RF
RF Augmented
SVM
SVM Augmented
0.0
0.2
0.4
0.6
0.8
1.0


Training
9P/Tempel
Testing
103P/Hartley
Precision
Recall
RF
RF Augmented
SVM
SVM Augmented
0.0
0.2
0.4
0.6
0.8
1.0

Results



Outright removing one mission is very detrimental, but adding extra classes to asses artifacts seem to work