National Aeronautics and Space Administration

An Astronomical SCMOS Image Previewer

Principal Investigator: Chengxing Zhai (398); Co-Investigators: Zaki Hasnain (312), Umaa Rebbapragada (398), Michael Shao (326), Navtej Saini (398), Russell Trahan (383)

Program: FY21 R&TD Innovative Spontaneous Concepts

Objectives

Our objective was to use machine learning (ML) to train an astronomical image previewer for data cubes taken by scientific CMOS (SCMOS) cameras observing the sky so that the previewer can discern between normal and anomalous behavior and classify the anomalous behavior into different categories, such as camera anomaly, a telescope tracking problem, clouds in the field, excessive background light level, an intrusion of bright transient objects like LEO satellites, etc.

Background

Modern SCMOS cameras offer imaging capability in large format with high frame rate yet very low read noise. Their usage in planetary sciences and astrophysics has been increasing, *e.g.* ISRO's Chandrayaan-1, NASA's Parker Solar Probe, and ESA's Solar Orbiter missions. The low read noise enables to use short exposure images to replace a long exposure image for high temporal sampling and robustness against any transient signals. For example, the synthetic tracking (ST) technique [1,2], a JPL developed technology, uses short exposure frames to avoid streaked images for detecting fast-moving asteroids/satellites. The frames are integrated in post-processing to track both moving objects and stars to gain sensitivity and provide accurate astrometry. Convinced by our successful demonstrations, the Air Force will deploy ST for the next generation Ground-Based Electro-Optical Deep Space Surveillance system. Because real data always contain anomalous cases, to ensure timely data processing, it is necessary to discriminate these cases in advance before they cause trouble in the data processing pipeline. SCMOS cameras can generate large volumes of data in short time, therefore it is not feasible to have human intervention to assess data quality. We propose to use machine learning to train an image previewer for determining anomalous frames/datasets.

Approach

We used astronomical data from a near-Earth object (NEO) survey using JPL's robotic telescopes at the Sierra Remote Observatory to study the anomalous behaviors in the images. The images were taken by three different CMOS cameras (1 ZWO ASI6200MM Pro and 2 QHY600M) with an exposure time of 5 seconds. Each dataset consists of 100 frames of images observing the same sky field. We focused 30 datasets (100 frames/dataset) with anomalous cases including excessive sky background, out-of-focus, unexpected telescope slewing, significant clouds as shown in Figure 1.

We develop a pipeline to analyze astronomical images using computer vision-based feature extraction and machine learning techniques to identify outliers (Figure 2). Concurrently, we built an interactive web-based dashboard (AstroView) to explore, analyze, communicate, and assist in labeling data (See screenshots in Figures 3-6). We hand-label four anomaly groups: 1) images with satellite trajectories, 2) images with telescope slewing, 3) images with cloud obstruction, 4) images with obstructions due to low elevation telescope pointing. These hand-labels are used to 1) assess the quality of the extracted features, 2) the ability of the clustering algorithm to differentia normal and outlier data, and similarly 3) the ability of the anomaly detection algorithm to detect these outlier images.

We extract five distinct sets (see 'Feature extraction' in Figure 2) of features in each image: 1) basic statistics to detect changes in the sky background, 2) histogram of the image intensities and corresponding entropy to also describe amount of intense stars and uniformity in the image, 3) focus features to detect clarity of the image, 4) blob detection features to estimate the number of visible stars, and 5) streak detection features to detect satellite trajectories and telescope slew lines. We also create a sixth set of features from the aforementioned features by normalizing each feature by the mean value over a whole dataset.

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California
www.nasa.gov

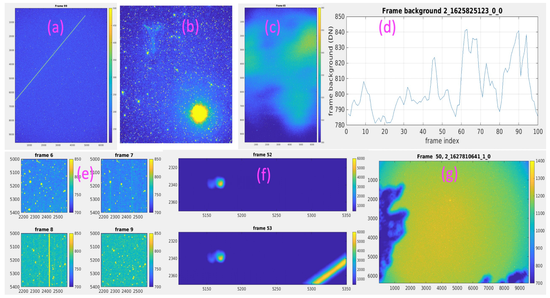

Figure 1. Examples of anomalous cases (a) a very bright satellite streak (b) ghost image of a bright star (e) clouds in field (d) excessive background variations from clouds (e) unexpected telescope slewing (f) out-of-focus images and a bright satellite (g) obstruction of tree leaves.

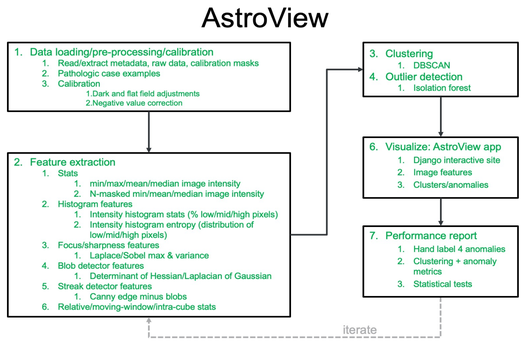

Figure 2. Flow-chart showing schematic of work done to extract features and find anomalies in astronomical images.

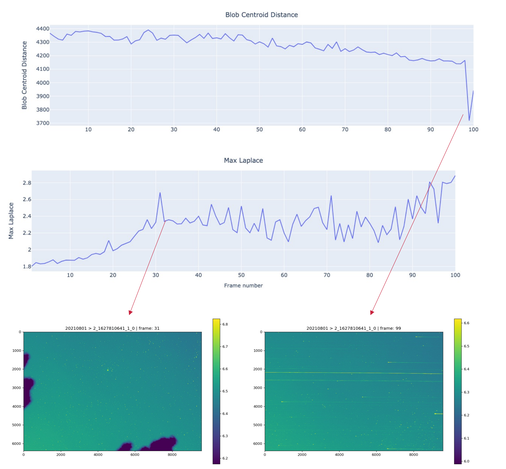

Figure 3. Distance to the centroid of blobs (top), maximum value of the Laplace filter (middle) for a dataset which contains obstruction anomalies as shown by image #31 (bottom left) and a slew anomaly as shown by image #99 (bottom right). The obstruction on the left edge of image #31 appears and starts moving across the field of view leading to changes in the Laplace filter values. Similarly the distribution of detected blobs shifts as the obstructions occupy more space.

PI/Task Mgr Contact
Email: Chengxing.Zhai@jpl.nasa.gov

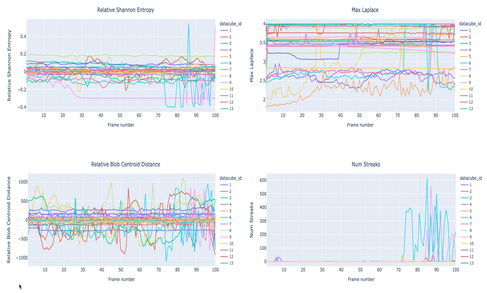

Figure 4. Examples of extracted features for 30 datasets (note legend is truncated as this is a screenshot of an interactive plot).

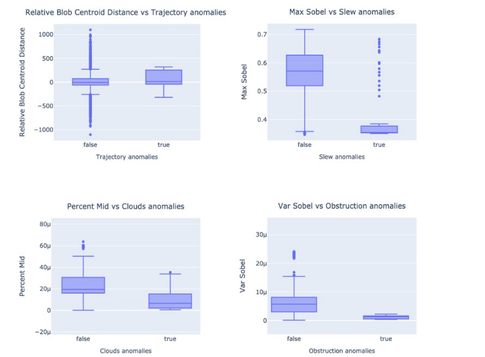

Figure 5. Comparison of extracted features to four hand-labeled anomalies for all 3000 frames in the dataset.

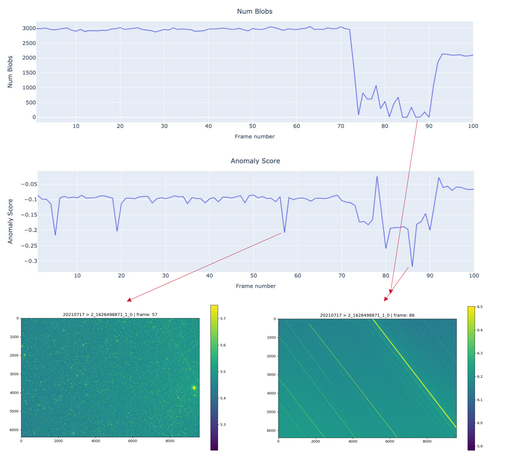

Figure 6. Number of blobs detected (top), anomaly scores from the Isolation Forest algorithm (middle) for a dataset which contains trajectory anomalies as shown by image #57 (bottom left) and slew anomalies from image #73-91, as shown by image #86 (bottom right). The number of blobs detected drops as the camera slews at image #73-91. More negative anomaly scores indicate presence of more anomalous features.

Results

Out of the total 3000 images, 1131 images were hand-labeled to contain any one of the four anomalies (237 trajectory anomalies, 75 slew anomalies, 614 cloud anomalies, 233 obstruction anomalies). We use Welch's t-test (Table 1) to evaluate whether the mean value of anomalous/non-anomalous frames is different for each feature. We found 8, 9, 13, and 14 features' mean values to be different between frames with/without trajectory, slew, cloud, and obstruction anomalies. We group all four hand-labeled anomalies two create a no-anomaly/any-anomaly set, and compare this classification to outliers found by the DBSCAN clustering algorithm and outliers detected by the Isolation Forest algorithm. DBSCAN clustering performed better in distinguishing the labeled anomalies (precision=0.74, recall=0.80, F1=0.77) as compared to Isolation forest (precision=0.55, recall=0.55, F1=0.49).

The blob detection tools developed successfully describe a nominal clear sky, and the streak detection tool is able to capture both slew and trajectory streaks. The combination of these features describes the imaging environment. These results suggest that our analytical pipeline comprising of computation tools and an interactive dashboard have made possible a future effort, which can successfully train machine learning classifiers to target specific anomalies.

Table 1. Welch's t-test results for difference in mean value of extracted features between anomalous and non-anomalous images. Four hand-labeled anomaly types: trajectory, slew, cloud, and obstruction due to surface features.

	Any anomaly		Trajectory		Slew		Cloud		Obstruction	
Feature	***t***	***p-value***	***t***	***p-value***	***t***	***p-value***	***t***	***p-value***	***t***	***p-value***
Max Laplace	-2.5	1.E-02	-1.4	2.E-01	-7.1	4.E-10	-5.4	7.E-08	5.2	5.E-07
Max Sobel	12.9	2.E-36	-14.0	3.E-35	11.9	4.E-19	8.7	2.E-17	10.3	3.E-21
Mean Masked	12.5	1.E-34	-3.8	2.E-04	3.4	1.E-03	9.4	3.E-20	17.0	3.E-50
Median Masked	6.6	6.E-11	0.4	7.E-01	-1.5	1.E-01	6.5	1.E-10	27.3	7.E-133
Num Streaks	-5.4	7.E-08	5.0	6.E-07	-6.7	3.E-09	5.4	7.E-08	5.4	7.E-08
Percent High	21.6	1.E-90	2.0	5.E-02	0.3	8.E-01	29.4	3.E-163	38.8	6.E-242
Percent Low	7.5	9.E-14	2.2	3.E-02	8.3	1.E-16	8.4	1.E-16	8.4	1.E-16
Percent Mid	25.5	2.E-121	1.5	1.E-01	0.7	5.E-01	33.3	4.E-200	41.3	3.E-258
Relative Blob Centroid Distance	-2.2	3.E-02	-6.9	2.E-11	1.1	3.E-01	-1.8	7.E-02	1.2	2.E-01
Relative Masked Mean	2.8	5.E-03	-5.0	1.E-06	2.8	6.E-03	-3.0	3.E-03	11.6	2.E-26
Relative Num Blobs	5.1	4.E-07	0.4	7.E-01	2.0	5.E-02	-1.5	1.E-01	30.2	8.E-175
Relative Shannon Entropy	3.6	3.E-04	-4.4	1.E-05	4.5	2.E-05	-2.9	4.E-03	10.5	3.E-22
Streak Pixels Percent	-4.6	4.E-06	4.4	1.E-05	-5.4	9.E-07	4.7	3.E-06	4.7	3.E-06
Var Laplace	18.2	9.E-70	-0.2	9.E-01	0.1	9.E-01	23.9	8.E-113	19.9	5.E-75
Var Sobel	24.4	2.E-117	1.2	2.E-01	1.1	3.E-01	24.9	2.E-119	49.4	0.E+00

Table 2. 11 extracted features of images are clustered using DBSCAN, and the reported outliers vs inliers are compared to hand-labeled no-anomaly vs any-anomaly groups for 3000 images. The outliers detected by DBSCAN. The clustering performed best in detecting obstruction and cloud anomalies but did not identify trajectory and slew anomalies well. Partly because the majority of those images are unaffected

Hand-label		Precision	Recall	F1	support
No anomaly	0	0.87	0.83	0.85	1869
Any anomaly	1	0.74	0.80	0.77	1131
Macro avg		0.80	0.81	0.81	3000
Weighted avg		0.82	0.82	0.82	3000

Significance/Benefits to JPL and NASA

We expect increasing usage of CMOS cameras for future Mar's explorations as well as other NASA missions. For example, Mars 2020 carried four CMOS engineering cameras. An image previewer that can identify anomalous images can help on-board/in-situ data processing to avoid potential complex consequences from processing abnormal data. We can also discard bad images in advance to reduce downlink data volume.

Acknowledgements

We would like to thank Dr. Seungwon Lee, group supervisor of 398K, for suggesting the spontaneous concept funding opportunity for this project.

References:

[1] M. Shao, B. Nemati, C. Zhai, et al, ApJ, 782, 1, (2014).
[2] C. Zhai, M. Shao, B. Nemati, et al, ApJ 792, 60 doi:10.1088/0004-637X/792/1/60, (2014).
[3] A. Mahabal, U. Rebbapragada, et al., Machine learning for the ZTF, PASP, 131, 997 (2019).

Clearance Number:
RPC/JPL Task Number: R21259