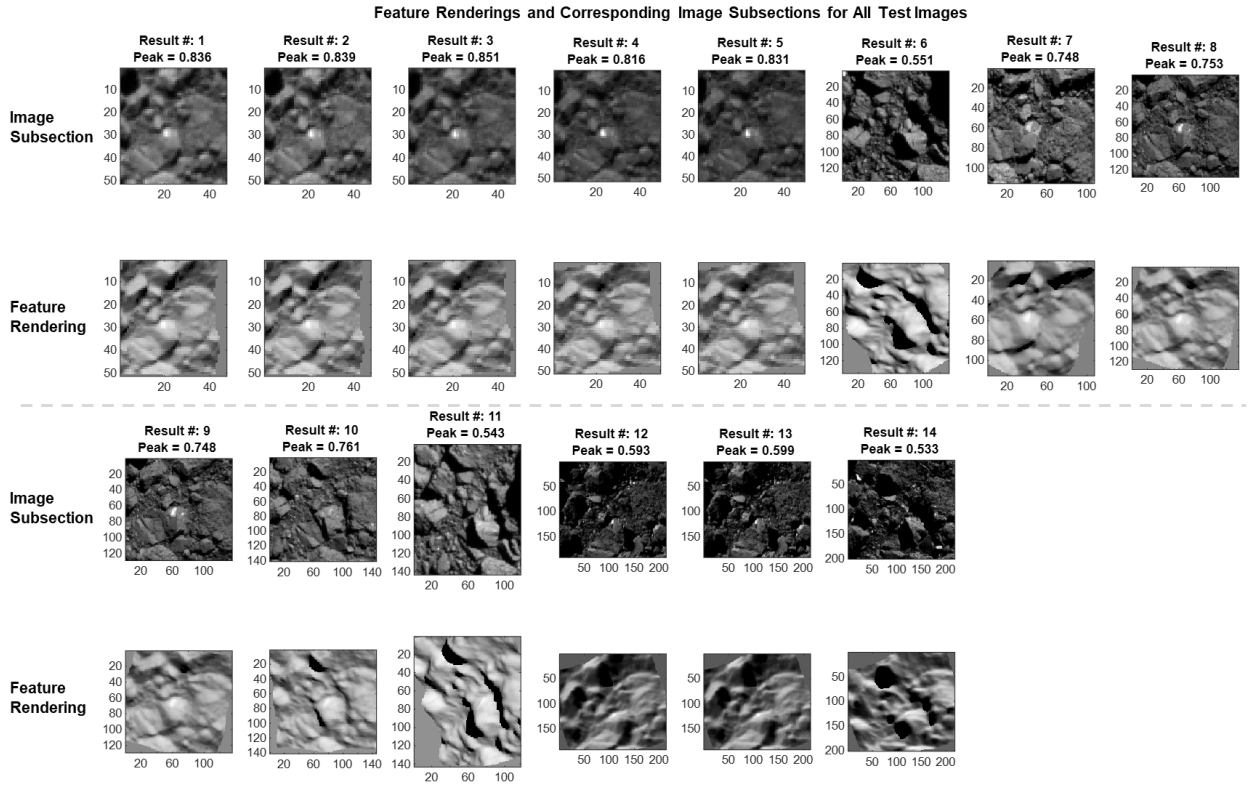
Correction: Development and Flight Performance of the Autonomous Navigation Feature Catalog for OSIRIS-REx Asteroid Sample Collection

Author(s) Name: Courtney Mario(1); Curtis Miller(2); Chris Norman(2); Ryan Olds(2); Eric Palmer(3); John Weirich(3); David Lorenz(4); Olivier Barnouin(5); Brent Bos(4); Bashar Rizk(6); Dante Lauretta(6)
Author(s) Affiliations: 1. Charles Stark Draper Laboratory Inc, Cambridge, MA, United States. 2. Lockheed Martin Space Systems Co, Denver, CO, United States. 3. Planetary Science Institute, Tucson, AZ, United States. 4. NASA Goddard Space Flight Center, Greenbelt, MD, United States. 5. Johns Hopkins University Applied Physics Laboratory, Laurel, MD, United States. 6. The University of Arizona Department of Planetary Sciences and Lunar and Planetary Laboratory, Tucson, AZ, United States.

Correction Notice

Correction 1
David Lorenz's affiliation should be:
KBR Inc., 7701 Greenbelt Road, Suite 400, Greenbelt, MD, 20771, USA

Correction 2a
Figure 9 has the image subsection and feature rendering rows mislabeled, potentially impacting interpretation of SPC model performance. This is the corrected image:



Correction 2b
Related to the above figure update, Figure 9 label should read as this instead:

Fig. 9 Feature renderings and corresponding image subsections for all 14 test images for Feature 270065. Results 1 to 8 are shown in the top half while 9 to 14 are in the bottom. For each result, the result is labeled with the primary peak correlation score, the top row shows the image subsection corresponding to the primary peak location, and the bottom row shows the feature rendering.

Development and Flight Performance of the Autonomous Navigation Feature Catalog for OSIRIS-REx Asteroid Sample Collection

Courtney Mario[1]
The Charles Stark Draper Laboratory, Inc., Cambridge, MA, 02139, USA

Curtis Miller[2], Chris Norman[2], Ryan Olds[3]
Lockheed Martin Space, Littleton, CO, 80125, USA

Eric Palmer[4], John Weirich[4]
Planetary Science Institute, Tucson, AZ, 85719, USA

David A. Lorenz[5]
NASA Goddard Spaceflight Center, Greenbelt, MD, 20771, USA

Olivier S. Barnouin[6]
The Johns Hopkins University Applied Physics Laboratory, Laurel, MD, 20723, USA

Michael G. Daly[7], Jeff A. Seabrook[8]
York University, ON, M3J 1P3, CAN

Brent J. Bos[9]
NASA Goddard Spaceflight Center, Greenbelt, MD, 20771, USA

Bashar Rizk[10], and Dante S. Lauretta[11]
University of Arizona, Tucson, AZ, 85721, USA

Abstract – The OSIRIS-REx spacecraft successfully collected a sample from the asteroid Bennu in October 2020, enabled by the vision navigation system Natural Feature Tracking (NFT). NFT autonomously provided state updates by matching features defined from Bennu shape model data to onboard camera images and allowed the spacecraft to touch down within 1 meter of the targeted location. This paper presents the development process and flight

[1] Principal Member of the Technical Staff, Perception and Autonomy.
[2] Guidance Navigation and Control Engineer.
[3] Guidance Navigation and Control Lead.
[4] Senior Scientist.
[5] OSIRIS-REx TAG Campaign Lead.
[6] OSIRIS-REx Altimetry Working Group Lead.
[7] OLA Instrument Scientist, The Centre for Research in Earth and Space Science.
[8] OLA Deputy Instrument Scientist, The Centre for Research in Earth and Space Science.
[9] Research Physicist, Goddard Space Flight Center.
[10] OCAMS Instrument Scientist, Lunar and Planetary Laboratory.
[11] OSIRIS-REx Principal Investigator, Lunar and Planetary Laboratory.

performance of the feature catalog used for navigation, including findings about terrain characteristics of robust features, validation methodologies despite limited test imagery, and trends between feature terrain content and final feature performance.

I. Introduction

On October 20, 2020, the OSIRIS-REx (Origins, Spectral Interpretation, Resource Identification, and Security–Regolith Explorer) successfully collected a sample from the asteroid Bennu [1]. To collect a sample, the spacecraft performed a Touch-and-Go (TAG) maneuver, contacting the surface with only the sampling head [2] before backing away. At the time of TAG, one-way light time to the spacecraft was over 18 minutes, requiring the spacecraft to be fully autonomous as it navigated to the asteroid surface. This was made possible in large part due to autonomous, camera-based navigation updates provided by Lockheed Martin Space's Natural Feature Tracking (NFT) algorithm, which enabled the spacecraft to touch down within 1 meter of the targeted location amidst an area surrounded by boulders and other spacecraft hazards [3][4].

During TAG, NFT operated autonomously onboard the spacecraft for over four hours, starting at the Orbit Departure Maneuver at approximately 700 meters above the surface and continuing through the last image at 27 meters just before surface contact. NFT provided spacecraft state updates by matching a catalog of known features to real-time images of the asteroid surface captured along this trajectory and successfully matched every feature, providing a final predicted TAG location that was only 3.5 cm from the actual surface contact point [3][4]. Figure 1 shows an image of the sample site on Bennu, overlaid with the hazard map used by NFT as well as the targeted, NFT-estimated, and actual contact locations, showing the hazardous surface with little room for error that NFT was able to deliver to. This success of NFT was due in large part to the design of the onboard feature catalog and the ability of feature data to correctly represent Bennu's terrain.

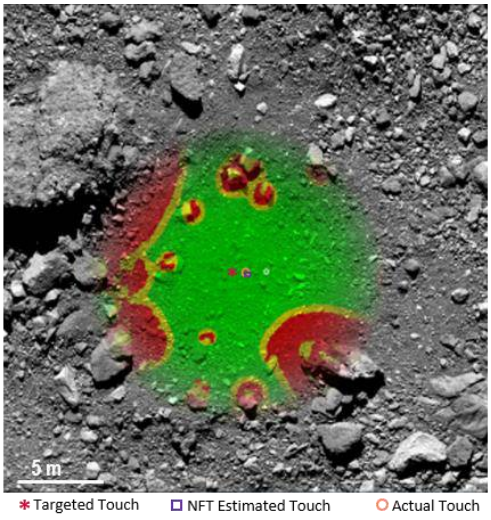


Fig. 1 OSIRIS-REx sample collection site, hazard map, and TAG performance. Image shows the sample site on Bennu where the OSIRIS-REx spacecraft successfully collected a sample from. The green, yellow, and red overlay represents the hazard map, with green regions determined to be safe for spacecraft contact. In the center is the targeted contact location (pink *), NFT's estimated contact location (purple square), and the actual spacecraft contact point (orange circle) – all within 1 meter of one another.

This paper presents the NFT feature catalog development process for the catalog used for TAG, including feature selection criteria for robust matching, feature validation approaches and results, and final feature catalog design and ordering. This paper also discusses feature validation challenges due to a lack of test images at the conditions expected for sample collection and how this was addressed through detailed feature performance analysis and spacecraft rehearsals. Lastly, this paper presents image correlation performance results for the NFT feature catalog during the sample collection as well as an analysis comparing trends in feature correlation performance to feature terrain content.

II. Background

A. Mission Overview

The OSIRIS-REx spacecraft launched in September 2016, arrived at Bennu in late 2018, performed orbit operations at Bennu until around March 2021, and will return the collected sample of Bennu's regolith to Earth in September 2023 [1]. Prior to arrival at Bennu, the team had an idea of Bennu's orbit, shape, mass, rotation state and other surface properties from ground-based observations that were used for mission planning and it was expected that available sample sites would be at least 50 m in diameter [5][6]. However, upon arrival at Bennu, the team discovered a rocky and rugged surface. As operations at Bennu progressed, the spacecraft flew a wide array of trajectories to collect data, learn more about Bennu, and identify potential landing sites. This culminated in selection of a primary sample site, Nightingale (Fig. 1), as well as a backup sample site, Osprey. Both of these sites were only 8-10 m in diameter and surrounded by potential spacecraft hazards, requiring the mission to upgrade NFT from a backup solution to the primary navigation method for TAG due to its increased precision over the originally planned lidar-based navigation method.

Payloads onboard the OSIRIS-REx spacecraft include scientific instruments designed for both science and navigation purposes. Those relevant to NFT include the TAGCAMS camera suite [7], the OCAMS camera suite [8], and the OLA scanning laser altimeter [9]. TAGCAMS has two cameras which were used for navigation – NavCam 1, which was used for ground-based optical navigation updates for navigation needs outside of TAG [10], and NavCam 2, which was used by NFT for navigation during TAG. OCAMS includes three cameras at a range of resolutions and was used to collect images to generate feature data for NFT and also provided the majority of the test images available for NFT testing prior to TAG. Lastly, OLA was used to build a high resolution model of Bennu and provided the TAG site Digital Terrain Model (DTM) used by NFT to predict surface contact [11].

Prior to TAG, the mission executed two TAG rehearsals: a Checkpoint Rehearsal and a Matchpoint Rehearsal. These were designed to fully validate NFT as well as the two trajectory correction maneuvers that would occur during TAG which were referred to as Checkpoint and Matchpoint. The Checkpoint Rehearsal took place in April 2020 and brought the spacecraft through the Checkpoint maneuver to an altitude of 86 m above the surface, while the Matchpoint Rehearsal took place in August 2020 and brought the spacecraft through both the Checkpoint and Matchpoint maneuvers to an altitude of 45 m above the surface. Both rehearsals went extremely well and proved that NFT and the spacecraft were ready for sample collection. In October 2020, NFT successfully guided the spacecraft for TAG, enabling a sample collection that appears to far exceed the required 60 g of regolith.

B. NFT Overview

NFT is a type of Terrain Relative Navigation (TRN) system as it provides absolute navigation updates by matching images to an existing terrain map. Unlike other TRN systems that match features to a map using pre-defined feature templates [12][13], feature descriptors [14], or surface patterns like craters [15][16][17], NFT is different in that it stores feature shape model data onboard so that it can render the expected appearance of each feature in every image using the predicted spacecraft state and sun angle of each image [18]. This approach provides the most accurate predicted feature appearance, increasing the likelihood of a valid feature match, but also increases onboard processing requirements. For OSIRIS-REx and due to NFT being added later as a backup solution [19], computational resources and timing constraints limited NFT to processing only five features per image. This meant little to no margin for outliers, or false feature matches, which made the feature selection, modeling processes, and catalog design critical to the overall success of NFT [20][21].

An NFT 'feature' here is defined as a patch of the surface and consists of a DTM and albedo layer, where the DTM defines the feature location and surface topography while the albedo defines how light reflects off the surface. These combined DTM and albedo products were produced by the OSIRIS-REx Altimetry Working Group (AltWG) team for NFT for use in the onboard feature catalog, leveraging shape modelling techniques that were already being applied for other OSIRIS-REx science and navigation needs.

During TAG, a total of 30 images were processed by NFT. For each image, NFT searched the onboard feature catalog to identify the first five features visible in that image. Then, for each feature, NFT used the data stored in the onboard feature catalog along with the predicted spacecraft position, spacecraft orientation, and sun vector to render the expected appearance of the feature. Next, each feature rendering was correlated to the onboard image using a search region defined from the predicted feature location in the image plus some margin to account for spacecraft uncertainty. The resulting correlation surface was then processed to extract correlation peaks and any peak scoring above 0.45 (potentially multiple per feature) was passed to the validation step. When all features in the image had been correlated, all peaks were processed by the validation step to remove any outliers and the resulting inliers were

used to compute a spacecraft state update [18][3][4]. An example of features processed by NFT for an image during TAG is shown in Fig. 2.

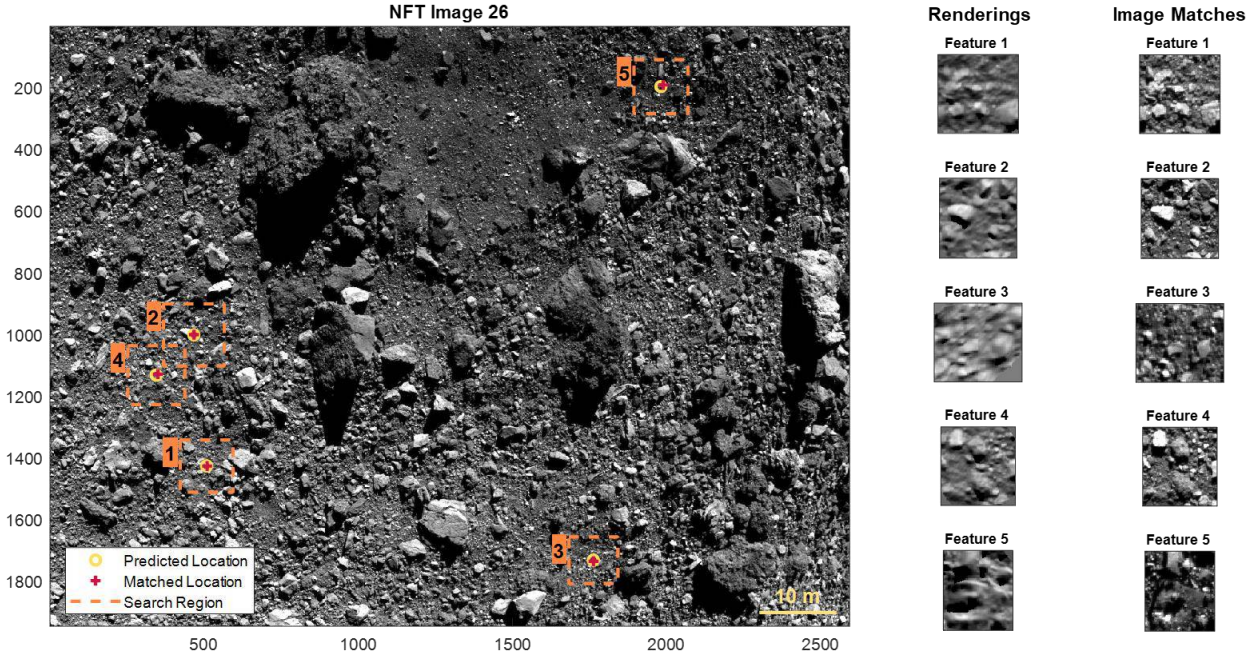


Fig. 2 NFT features for image 26 during TAG. The left side shows the 26th NFT image of the TAG trajectory, overlaid with information about the 5 features processed by NFT. For each feature, the orange box is the correlation search region, the yellow 'o' is the predicted feature location, and the red '+' is the matched location. The middle column shows the feature renderings using DTM and albedo data, while the right column shows the corresponding matched regions from the image.

C. Bennu Shape Model Generation

For the OSIRIS-REx mission, AltWG utilized two different methods to produce shape models of Bennu: Stereophotoclinometry (SPC) and a process leveraging data from OLA. SPC solves for topography and albedo simultaneously from images at varying emission angles, incidence angles, and resolutions, leveraging both stereo imagery techniques and photoclinometry [22][23] and was used by the Flight Dynamics Team for ground-based navigation updates [10]. Alternatively, OLA uses scanning lidar measurements to provide topography data [24][25] and provided high-resolution information to assess candidate TAG sites along with the DTM used by NFT to predict surface contact [4]. Both methods were evaluated for NFT feature generation during a detailed ground testing campaign prior to Bennu arrival [26] but due to NFT's dependence on albedo data, the presence of strong albedo features on Bennu [27][28], and challenges incorporating albedo into OLA features, it was decided to use SPC for NFT feature generation.

Unlike other TRN approaches used for lunar or Mars missions, a "map" of Bennu did not exist prior to the mission as the only available Bennu data was from Earth-based observations. This meant that the OSIRIS-REx mission needed to plan for operations at Bennu that enabled collecting OCAMS images and OLA scans to support both SPC and OLA shape model generation. Both models were continually refined as data at increasingly higher resolutions were collected during asteroid operations. Figure 3 shows the timeline of asteroid observation phases at Bennu used for this data collection. Significant effort went into planning spacecraft trajectories during these phases [1] to ensure adequate data was available to build NFT features at the fidelity needed for robust correlation. This timeline and data availability ultimately drove the NFT feature generation and testing schedules as well.

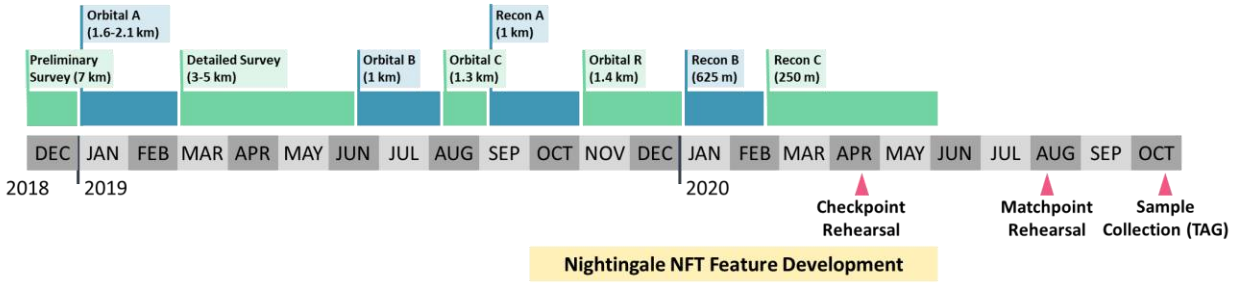


Fig. 3 OSIRIS-REx Observation Timeline. The green and blue boxes name the different asteroid observation phases, with approximate distances to Bennu's surface given in parentheses. For comparison, NFT's feature development timeline for the Nightingale site is shown by the yellow box on the bottom.

D. Correlation Evaluation Criteria and Terminology

With correlation, a feature patch is matched against an image search region and a correlation surface is produced which contains a score of how well the feature patch matches the image at ever pixel within the search region. Scores can range from -1 to 1, where 1 is a perfect match and -1 is an exact inverse match. Ideally the result is a surface with a single high scoring "primary" peak representing the correct feature location, without any other high scoring "secondary" peaks at incorrect feature locations nearby, and with a small pixel error between the primary peak location and the true feature location. However, it's possible to see multiple high scoring peaks or large pixel errors due to differences between the appearance of the feature in the image and the data contained in the patch.

Since NFT uses a method of normalized cross correlation to match features to onboard images [18], the same approach was used to evaluate feature shape model data. In addition, since NFT applies a correlation score threshold to the correlation surface and sends all peaks scoring above that threshold to the validate step, correlation score became an important metric for feature evaluation. Pixel error between the feature's true location and its detected location in the image was another important metric to ensure feature location errors did not introduce errors to the state update. During feature evaluation, features were considered passing if the primary peak scored above 0.45, all secondary peaks scored less than 0.4, and the pixel error of the primary peak was less than 6 pixels, which was determined from analysis as the requirement needed to ensure NFT could meet TAG navigation requirements.

Image conditions were the other important aspect of feature evaluation in order to fully characterize feature performance. These conditions included ground sample distance (GSD), sun incidence angle, sun azimuth angle, camera emission angle, and camera azimuth angle. GSD defines the ground distance between the centers of two consecutive pixels and is used in this paper as distance/pixel (i.e. m/pixel). The sun angles define the angles between the sun and the surface at the center of the feature, where incidence is the angle between the sun and the surface normal while azimuth is the angle between the sun and the y-axis of the feature surface plane. Similarly, the camera angles define the angles between the camera boresight and the surface at the center of the feature, where emission angle is the angle between the camera boresight and the surface normal and azimuth is the angle between the camera boresight and y-axis of the feature surface plane. The use of "image conditions" throughout this paper refers to the collection of these five terms.

III. Feature Catalog Development Process

The flow diagram in Fig. 4 presents the development process to build the feature catalog that enabled NFT's successful autonomous navigation for TAG. This section outlines this approach, describing methods for the feature selection process performed by the NFT team (III.A), the feature generation process performed by AltWG (III.B), the feature evaluation process performed by the NFT team (III.C), and the final processing steps performed by both teams to produce the final feature catalog (III.D). Later, Section V provides detailed results from this approach as it was applied to produce the TAG catalog. While this paper focuses on the efforts to produce a TAG catalog for Nightingale, the same process was also completed for the backup site Osprey to ensure that the team could transition to a second TAG attempt if needed.

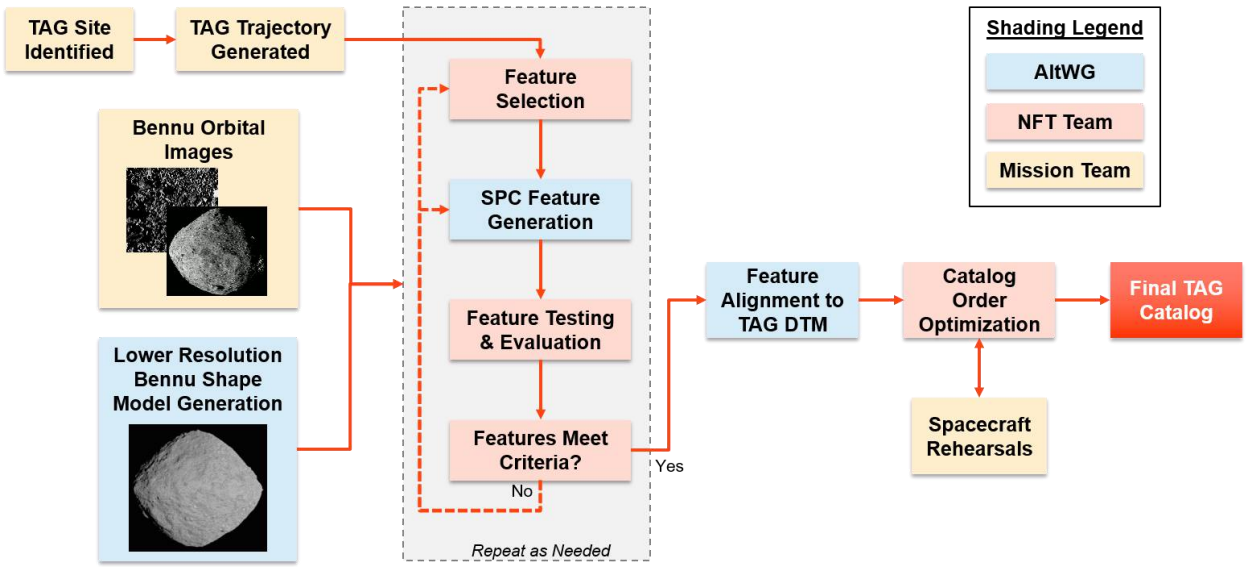


Fig. 4 NFT feature catalog generation process.

A. Feature Selection

The NFT feature catalog needed to be designed so that it would provide coverage for all NFT images in the TAG trajectory, including up to 3-sigma trajectory dispersions. To ensure that 5 features were visible in every image, it was expected that 100 to 150 features would be needed to cover the entire trajectory. In addition, feature data needed to support images at 20 cm/pixel down to as small as 0.3 cm/pixel and cover areas large enough so that the rendered feature size in the intended image would be approximately 100 x 100 pixels. Due to mission timeline and imaging constraints, it was not practical to produce a global SPC model of Bennu at the smallest GSD needed for NFT features (0.5 cm/pixel). Instead, AltWG generated an SPC model that was 14 cm/pixel globally and 5 cm/pixel near the TAG sites, and then focused on generating higher resolution data at locations where the NFT team determined features were needed to support the TAG trajectory. This also meant that the NFT team could only use available OCAMS images for feature selection. And due to the time-intensive process for AltWG to generate features using SPC, the mission timeline could not afford several iterations of feature generation if the NFT team selected feature terrain that could not be properly modeled by SPC.

In order to select features with terrain that SPC could model well and that would be robust for correlation, the team leveraged lessons learned from years of testing with synthetic shape models [9][26] and three NFT checkouts in 2019 [21][29]. This testing determined that SPC often smooths the edges of terrain, impacting the shapes of shadows rendered by NFT and potentially impacting correlation performance. Testing also found that SPC struggled with taller terrain features, which was often challenging given Bennu's rugged terrain. However, SPC was able to model albedo variations well, which was especially valuable since Bennu had greater albedo variation than originally expected [28]. This understanding of the strengths and weaknesses of SPC became important for NFT feature selection.

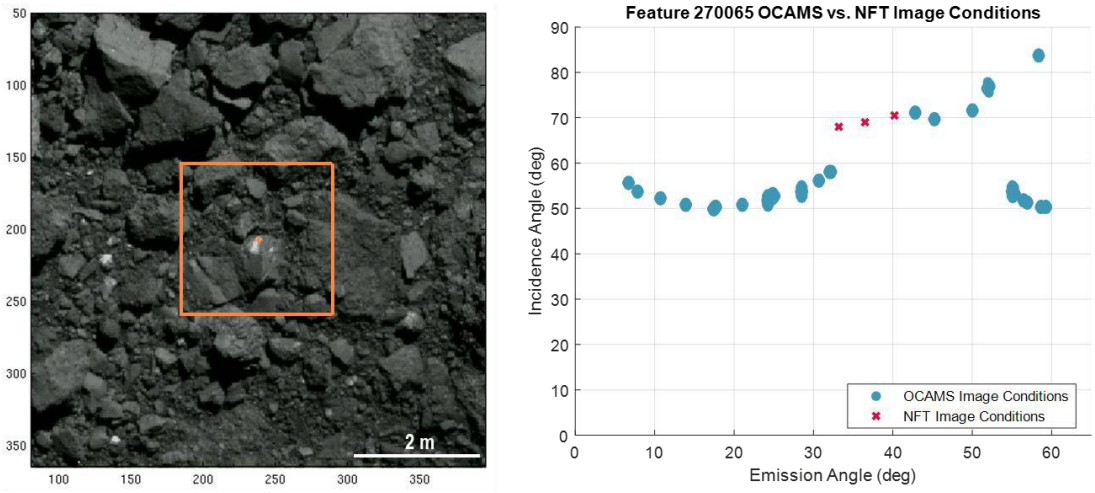
The general feature selection process was to first identify the surface coverage area of the nominal TAG trajectory and 3-sigma dispersions along with the required feature GSDs. To make processing easier for AltWG, the NFT team divided up the feature selection process into bundles based on the feature GSD needed to support ranges of NFT images. This allowed AltWG to focus on small portions of the feature development at a time, starting with the largest feature GSDs and progressing towards smaller feature GSDs as the mission collected more images for SPC. It also enabled the NFT team to start testing as soon as a bundle was ready, allowing the team to refine the testing and evaluation process over time as well as learn more about SPC performance at Bennu and incorporate additional lessons into feature selection for later bundles. The feature bundles, target image GSDs, and number of features allocated per bundle are presented in Table 1. While most bundles had 15 to 20 features, the first bundle (largest GSDs), had 28 features since these were easier for AltWG to generate using the 14 cm/pixel global model as a starting point. Also, even though NFT only processed 30 images during TAG, the feature development process included 31 NFT images to account for a small set of cases (65 of 1000 Monte Carlo runs) where NFT processed a 31st image depending on the trajectory flown.

Table 1 NFT Feature Selection Bundles

Feature Bundle	NFT Images Targeted	GSD Range of Target Images (cm/pix)	Estimated # of Features
1	1 to 11	6.9 to 16.6	28
2	12 to 17	3.1 to 5.8	19
3	18 to 24, Part 1	2.8 to 3.1	17
4	18 to 24, Part 2		17
5	25 to 27, Part 1	1.6 to 2.6	17
6	25 to 27, Part 2		17
7	28 to 29	0.8 to 1.1	15
8	30 to 31	0.3 to 0.5	15

To select features for each bundle, the NFT team analyzed all OCAMS images that covered the NFT image footprints within that bundle. The goal was to identify patches of terrain that the team expected SPC could correctly model and were at the size and GSD needed to support the intended NFT image. Once a feature patch was identified for potential consideration, the NFT team analyzed it in several different ways. The first was to correlate the patch with all other available OCAMS images of that feature. While this did not test SPC's correlation performance, it did allow the NFT team to determine 1) how well that image patch correlated at a range of image conditions, 2) whether any nearby terrain was similar and therefore increased the likelihood of false matches, and 3) the range of available OCAMS image conditions available. The OCAMS image conditions were an especially important aspect of feature selection since they provided insight into SPC model quality, where a wider range of image conditions improved SPC performance [11]. Also, since these OCAMS images would be the only images available for NFT feature testing, they also let the NFT team assess how closely the test images would match expected TAG image conditions.

Figure 5 shows an example of some of the information used by the NFT team during this process. In the image of the candidate feature on the left, while the terrain selected contains rocks that may be smoothed during SPC processing, it is relatively flat compared to the surrounding boulders and the terrain and albedo pattern is distinct, making it a good candidate feature. In the plot on the right comparing OCAMS image conditions to NFT image conditions during TAG, while no OCAMS images exactly overlap the NFT image conditions, they do span both sides. This meant that for testing, if the feature correlated well across all images, it would be expected that the feature would also correlate well for NFT images.



a) Feature Identified in OCAMS image

b) Feature Image Conditions

Fig. 5 Example feature selection results. (a) Feature 270065 as identified in an OCAMS image during feature selection. (b) A comparison of OCAMS image conditions (blue dot) for Feature 270065 available for both feature generation with SPC and feature testing against the expected image conditions for Feature 270065 during TAG (red 'x').

Lastly, the NFT team ran a Monte Carlo simulation of the TAG trajectory to evaluate how much coverage of the 3-sigma dispersed trajectories each feature would provide as well as the expected image conditions for each feature. Features that had either camera emission or sun incidence angles that were too large were thrown out, since both cases were found to lower correlation scores. This process was repeated until the NFT team had identified a set of features with terrain that was expected to work well for correlation, provided adequate coverage of all dispersed trajectories to

ensure 5 features would be available for correlation for each image, and were positioned so that they provided adequate spread across each image.

B. Feature Generation

When feature selection was complete for each bundle, the NFT team sent the necessary feature data to AltWG. To avoid any potential mix-ups in feature locations or sizes, this data included a distinct feature ID, names of OCAMS reference images along with pixel coordinates where the feature was located, and images of the feature as identified by the NFT team (i.e. Fig. 5a). In addition, data also included the feature location in both Asteroid-Centered-Fixed (ACF) and latitude/longitude coordinates, the desired feature GSD, and the physical size of the feature size in meters.

AltWG then began refining data from the global shape model in these locations to build features at the specs needed for NFT, leveraging available OCAMS images to iterate and improve the SPC model at each location. Prior to delivering feature models to the NFT team, AltWG performed their own evaluation of the features, which included performing correlation tests as well as other manual and automated tests to identify any drastic model errors. This process took on the order of days to weeks, with the exact time dependent on the quantity of the features and the desired GSD.

C. Feature Evaluation

Once the feature DTM and albedo data were ready, the NFT team tested the features against all available OCAMS images to evaluate correlation performance. This process involved analyzing feature renderings and correlation performance on a feature by feature basis, carefully evaluating SPC's ability to represent terrain and feature performance at specific image conditions in an attempt to predict each feature's ability to correlate well for TAG image conditions. Feature correlation performance was evaluated primarily based on correlation score, to ensure features would pass NFT's correlation score threshold check, as well as pixel error needed to meet navigation requirements. More information about this process and example results are provided in Section IV.

Any features that did not meet correlation requirements were then reviewed by the NFT and AltWG teams together to assess whether AltWG thought model quality could be improved with additional processing or if the NFT team needed to select a different feature. The feature selection, generation, and evaluation steps would then be repeated as needed until the NFT team confirmed that they had sufficient feature coverage to meet their navigation requirements for TAG.

D. Final Feature Catalog Preparations

After the full set of NFT features were generated and passed correlation evaluation, the NFT and AltWG teams needed to perform the final steps to finalize the TAG catalog. The first part of this process required aligning the NFT features with the OLA DTM to improve feature location accuracy and ensure consistency with the DTM used by NFT to estimate surface contact. This process was performed by AltWG and used an iterative closest point method to find the position of the NFT feature that minimized the difference in topography between the NFT DTM and the OLA DTM. This became challenging for NFT features with the smallest GSDs since the OLA DTM was at 4.8 cm/pixel and NFT feature GSDs were as small as 0.5 cm/pixel. For these features, AltWG applied a fixed translation based on the known offset between SPC and OLA DTMs from prior characterization of the two models. In a few cases, features still had large errors after this fixed translation was applied so were manually adjusted further.

The last step was to optimize the feature catalog order. For the Checkpoint Rehearsal, it was determined that the default order of the features as they were selected was sufficient and did not need to be further optimized. However, the catalog order was critical for the post-Checkpoint features for images 25-30. Since NFT could only process five features per image, it was important to ensure that these features were 1) dispersed across the image to improve observability and 2) would be robust for correlation. Therefore, this process required optimizing feature order across multiple variables including the range of images each feature would be used, feature position in each image, expected correlation performance for the range of image conditions, and 3-sigma trajectory dispersions. Lastly, the Checkpoint and Matchpoint rehearsals provided an opportunity to validate feature performance under TAG image conditions using the NavCam 2 camera and was especially useful to finalize the feature catalog order.

IV. Feature Validation Challenges and Mitigation Strategies

There were a few key challenges for feature development and validation testing. The first was due to the fact that by its nature, SPC produces models that match the terrain reflected in the images processed. When fewer image condition variations are used, this can lead to SPC models being biased towards topography and/or albedo solutions that do not correctly capture the terrain's appearance at all viewing angles and lighting conditions. Related to this

challenge was the fact that only OCAMS images were available for NFT feature testing prior to TAG, which were the same images processed by SPC to generate NFT feature data. On top of that, very few of the OCAMS image conditions matched the image conditions expected to be processed by NFT which employed NavCam 2 during TAG. As a result, there was a concern that validating NFT features with OCAMS images could result in features that correlate well in OCAMS images but fail to correlate for NFT during TAG.

The NFT team addressed these challenges in a few different ways. The first was to add margin to what was considered an acceptable correlation score for feature testing in OCAMS images, to ensure that even if correlation scores decreased for TAG image conditions, they would still pass NFT's correlation score threshold. While the correlation score threshold used by NFT was 0.45, the team aimed to have features score 0.6 or higher in all available test images. The other approach was to rigorously evaluate NFT feature performance in OCAMS images, studying feature renderings at varying image conditions to understand potential SPC modeling errors, evaluating correlation performance as a function of image conditions, and combining all results with a comparison to expected TAG image conditions to try to predict correlation performance at TAG. In cases where it was not possible to improve feature shape model data so that the feature would score greater than 0.6 in all test images, the team would analyze feature renderings and image conditions to determine why the feature had a lower score. If the team determined that either the lower scoring image conditions matched expected TAG conditions, or that the TAG image conditions would not experience whatever effect was lowering the correlation score (for example, overexposed albedo), then the team still considered a feature to be passing. In addition, all of these feature correlation performance considerations were tracked and used to optimally order the feature catalog.

Figures 6 to 9 provide an example of this feature validation process for Feature 270065, the same feature shown in Fig. 5. The primary figure used for analysis is shown in Fig. 6a, where the goal was to show that the primary peak (Peak 1) always scored at least above the 0.45 correlation score threshold, though ideally above 0.6, with pixel errors less than 6 pixels, with all secondary peaks (Peak 2, 3, etc.) scoring below well below the 0.45 correlation score threshold. This was done to ensure that only correct matches would be processed by NFT to compute the state update. For the results shown for this feature, larger pixel errors (50 to 100 pixels) for some results were due to small errors in spacecraft kernels impacting the accuracy of the projected feature location in the image. These results were for Recon C images, which were taken at the lowest observation altitudes where the image GSD in many cases was less than 1cm. So while the pixel errors were large, Fig. 6b shows that these large pixel errors correspond to approximately only 1 m error. In these cases, the NFT team also visually inspected the images and feature match locations to confirm these were valid matches, despite the larger pixel errors.

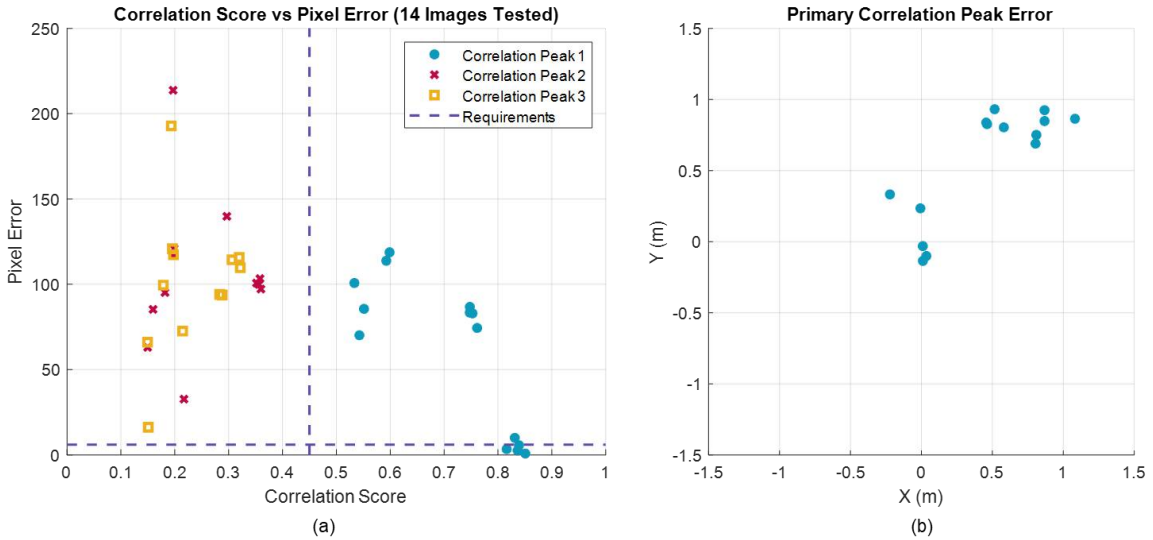


Fig. 6 Correlation Performance for Feature 270065 in all available OCAMS test images. (a) 3 highest scoring correlation peaks for each test image, where "Peak 1" (blue dot) is the highest, "Peak 2" (red 'x') is the second highest, and "Peak 3" (yellow square) is the third highest. The requirements used for evaluation, a minimum Peak 1 correlation score of 0.45 and pixel error less than 6 pixels, are shown by the dotted purple lines. Pixel error is computed as the difference in pixels between the predicted feature location in the image, computed using available spacecraft kernels, and the matched location. (b) Lateral errors in meters associated with each of the primary peak results (blue dots) in the plot on the left.

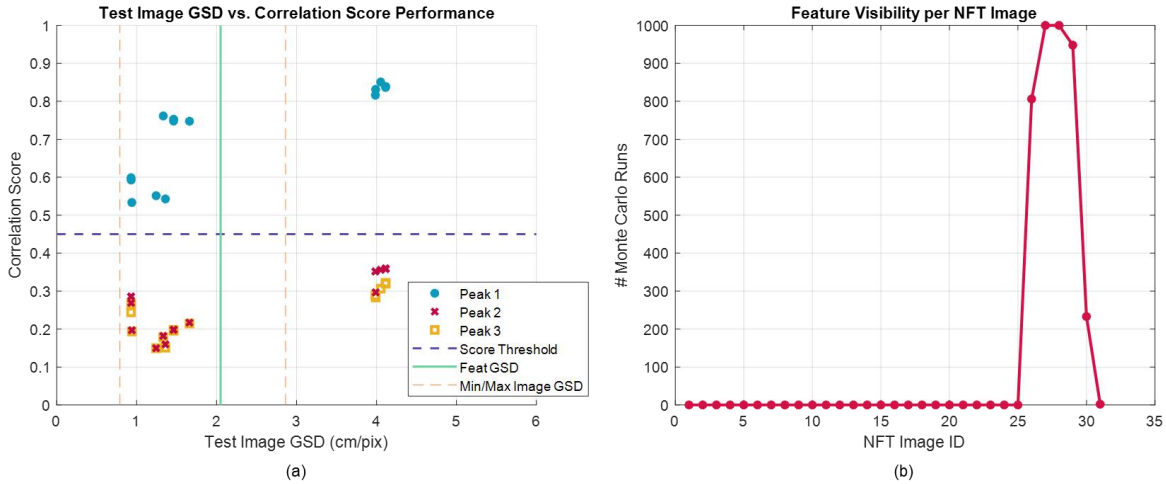


Fig. 7 Correlation score performance by GSD and Monte Carlo coverage for Feature 270065. (a) Same results as shown in Fig. 6a, except here correlation score is plotted as a function of test image GSD. The GSD of the feature is shown as the green line while the orange dotted lines represent the range of NFT image GSDs this feature could be processed for. (b) Visibility of feature in NFT images over 1000 Monte Carlo runs.

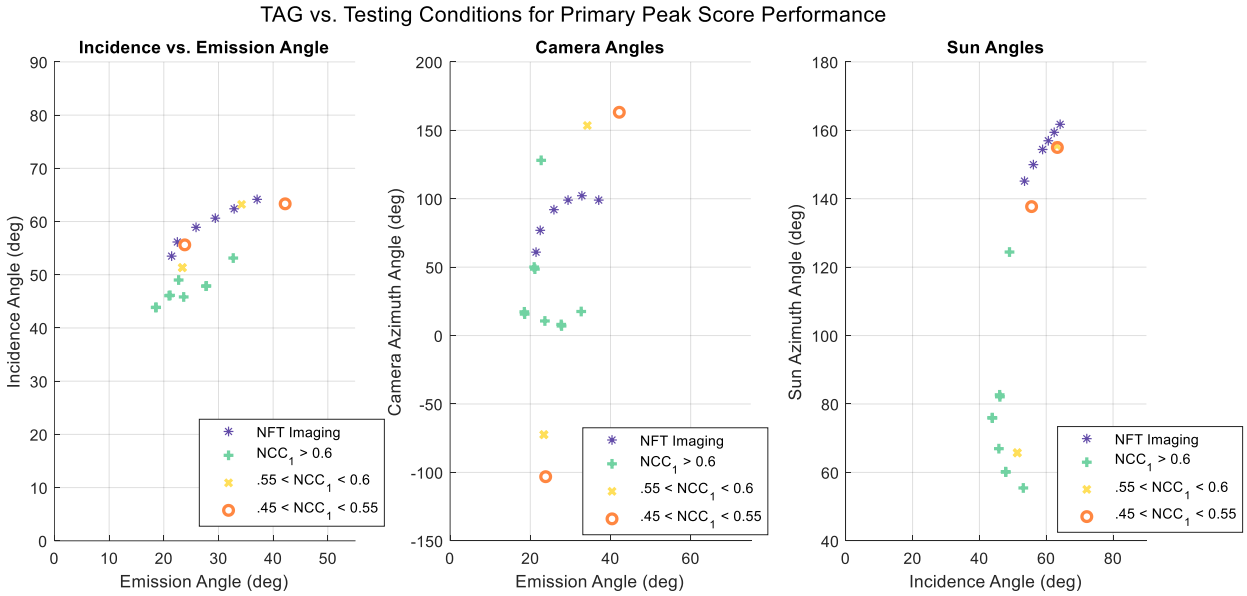


Fig. 8 Feature 270065 correlation score performance as a function of image conditions for test images and expected NFT image conditions. In all plots, the purple '*' represents the expected NFT image conditions for the nominal TAG trajectory while all other points represent test image correlation results. Test image conditions where the primary peak scored above 0.6 are shown as green '+', between 0.55 and 0.6 as a yellow 'x', and between 0.45 and 0.55 as an orange 'o'.

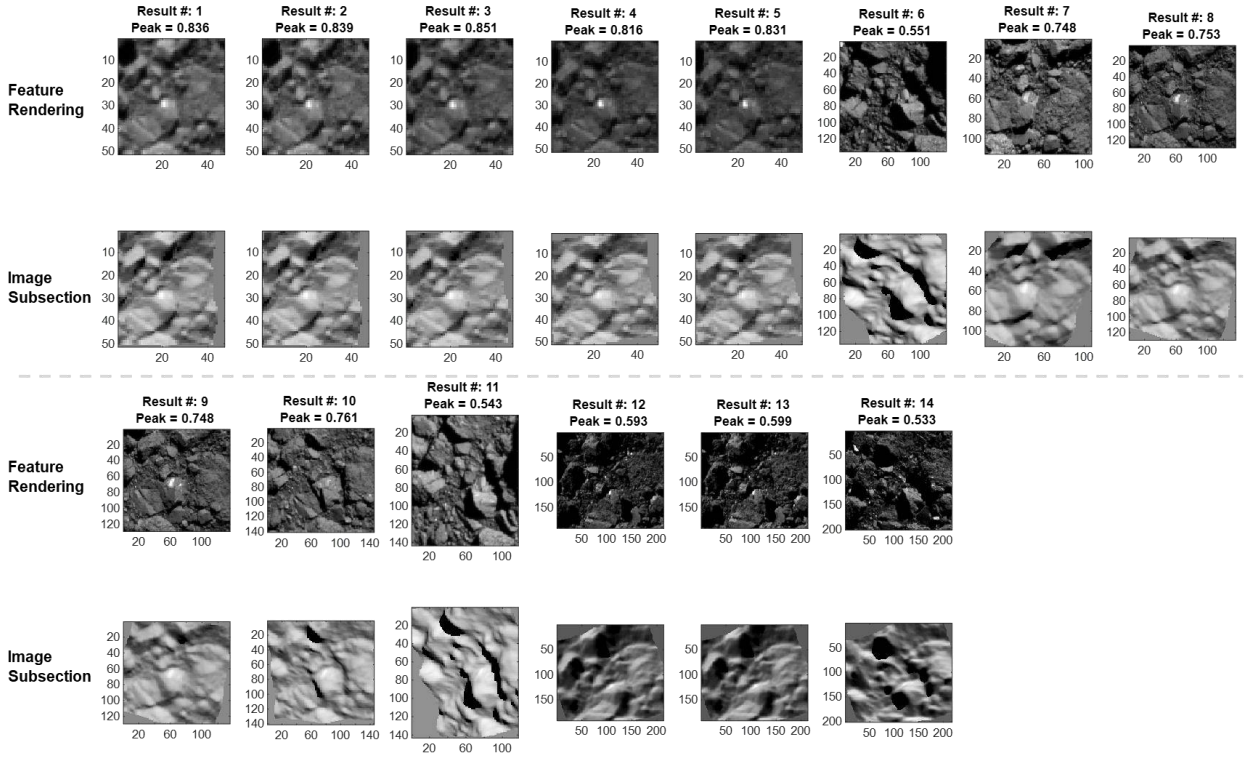


Fig. 9 Feature renderings and corresponding image subsections for all 14 test images for Feature 270065. Results 1 to 8 are shown in the top half while 9 to 14 are in the bottom. For each result, the result is labeled with the primary peak correlation score and the top row shows the feature rendering while the bottom row shows the image subsection corresponding to the highest scoring correlation peak.

For this feature, these results show that the feature scored greater than 0.7 for most test images, though there were a handful of lower scoring results that did not meet the 0.6 target score threshold. Further analysis determined that one of these lower scoring results occurred at a larger incidence angle where a rock outside of the feature cast a shadow across feature in camera image (see result #6 in Fig. 9, where shadow is in top right corner of image subsection), lowering the correlation score. Analysis showed that this would occur at sun angles similar to TAG for NFT images 28-31 so it was decided to only use this feature for images 26-27. The other low scores (see result #14 in Fig. 9) were due to both GSD and sun angle. For GSD, the feature was being used outside of a reasonable range for the underlying feature data, where the feature was built at 2 cm/pixel but was being tested on an image with a GSD at half that size. This was ultimately not a concern as the previous decision to not use the feature beyond image 27 meant it would not encounter these image GSDs during TAG. In addition, the rendering result in Fig. 9 showed that the SPC model had some terrain errors that caused shadows to be rendered incorrectly at this sun angle. Despite these errors, the feature was still able to correlate due to the overall shadow pattern being correctly represented. This result also confirmed that the shadow cast from the nearby rock would not be a concern for NFT images 26 and 27. Based on this feature's rendering performance at GSDs similar to images 26-27, and the visual inspection of the appearance of shadows and albedo in the feature rendering at image conditions similar to 26-27, the NFT team felt comfortable using this feature in the TAG catalog despite not all of the results scoring above 0.6. During TAG, this feature was processed for image 27 and scored 0.77, confirming this assessment was accurate.

This detailed review of feature correlation results and rendering was repeated for every feature and every re-delivery of reprocessed features, which made for a very manual and time-intensive process. While feature testing was automated as much as possible, it was not possible to automate this evaluation process since no test images matched expected TAG conditions for all parameters (GSD, incidence, emission, sun azimuth, camera azimuth). Adding margin to these parameters to find test images that matched at least a subset of the parameters quickly complicated the evaluation to where it was hard to have confidence in feature performance without a visual, per feature assessment like the one described here for Feature 270065. While this was tedious, the team felt confident that this approach would yield a robust catalog for TAG.

Lastly, the two spacecraft rehearsals provided valuable data to evaluate NFT feature performance using NavCam 2 under the expected TAG image conditions. This also provided an opportunity to compare predicted feature performance from the evaluation process using OCAMS images to their actual TAG performance. Section V presents the results from these rehearsals and how they were used to improve the NFT catalog design.

V. TAG Feature Catalog Development Details

The final catalog for TAG had 104 features with resolutions ranging from 15.6 to 0.84 cm/pixel. A total of 124 were selected by the NFT team and generated by AltWG using SPC. Of these 124 features, 13 were reworked by AltWG to improve the model quality while 5 were deemed unfixable and alternative features were selected instead. A total of 20 features not used because they did not meet correlation performance requirements. Table 2 shows these statistics and how they break down per grouping of NFT images.

Table 2 TAG Catalog Feature Development Statistics

Image Range	Features Generated	Features Reworked	Features Reselected	Features Not Used	Features Used in TAG Catalog
Image 1 to 11	18	2	2	2	16
Image 12 to 17	20	0	1	5	15
Image 18 to 24	33	1	1	4	29
Image 25 to 27	25	2	1	1	24
Image 28 to 29	15	2	0	4	11
Image 30 to 31	13	6	0	4	9
Total	**124**	**13**	**5**	**20**	**104**

A. OLA DTM Alignment Results

As discussed in Section III.D as one of the last steps to finalize the catalog, Fig. 10 shows the results of aligning all 104 features to the OLA DTM. The total amount features were translated varied from approximately 0.25 m to 0.8 m. The RMS error between the topography of the feature DTM and OLA DTM decreased for all features after the alignment too, with most features having a final RMS error less than 0.1 m. These are well within the uncertainty of 15 cm known to exist for the global shape solution of Bennu [30].

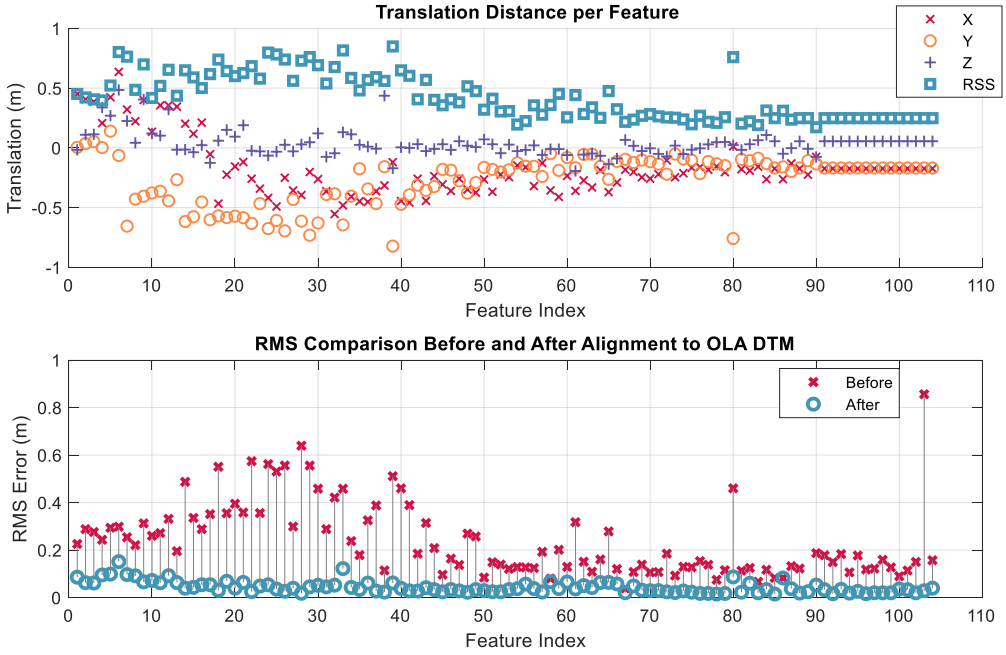


Fig. 10 Alignment of NFT Features to OLA DTM. The top plot shows the total amount each feature was were translated (with X, Y, and Z given in ACF) while the bottom plot shows a comparison of RMS error for each feature before and after they were aligned to the OLA DTM. Here, RMS error is computed as the root-mean-square of the 90th percentile of difference values between points in the feature DTM point OLA DTM. Feature index represents the index of the feature in the final TAG catalog.

B. TAG Rehearsals Correlation Performance

Both the Checkpoint and Matchpoint rehearsals provided valuable data for feature validation for NFT. Of the 31 NFT images that could be processed, the Checkpoint Rehearsal processed images 1 to 26, while the Matchpoint Rehearsal proceed further through image 28. After the rehearsals, the NFT team analyzed feature correlation performance of the features processed during the rehearsals and also correlated all visible features in all images to better understand performance for all features. The team also compared performance in rehearsal images at the expected TAG conditions to test images to determine how well the test image performance predicted performance for TAG.

Figures 11 to 16 show the analysis results from processing catalog features with the rehearsal images as well as comparing rehearsal performance to test image performance. Figure 11 shows that all features tested (using all available features, not only those processed during the rehearsals) scored above NFT's 0.45 score threshold, with the majority scoring above 0.7. In addition, all secondary peaks scored below 0.4 showing no images were at risk of processing false matches. While some pixel errors were large (up to ~25 pixels), this was expected since the errors were computed using NFT's predicted state for each image as no truth data was available and additional analysis confirmed that all primary peak represented correct feature locations. Figure 12 shows the feature performance based on the catalog design and was useful to finalize the catalog order that should be used for TAG. In some cases, poorer performing features were moved to the end of the catalog.

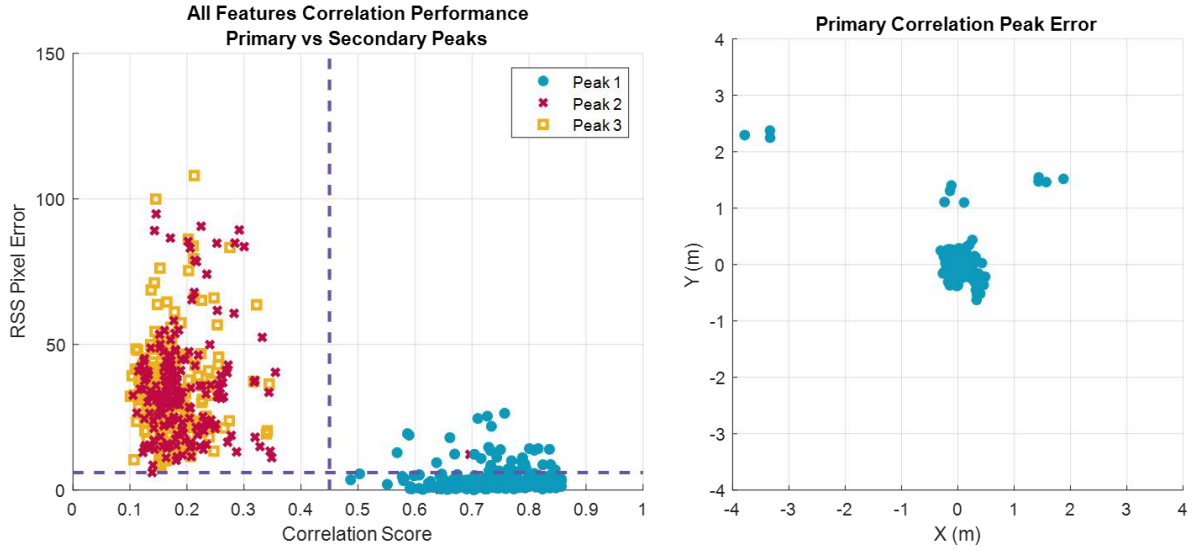


Fig. 11 Feature correlation performance for all catalog features using all images from the Checkpoint and Matchpoint Rehearsals. Pixel error in the plot in the left is not a true pixel error measurement since it is computed using NFT's predicted state for each image, which had a larger error for some images. The plot on the right shows the corresponding physical distance error of each blue point in the left plot.

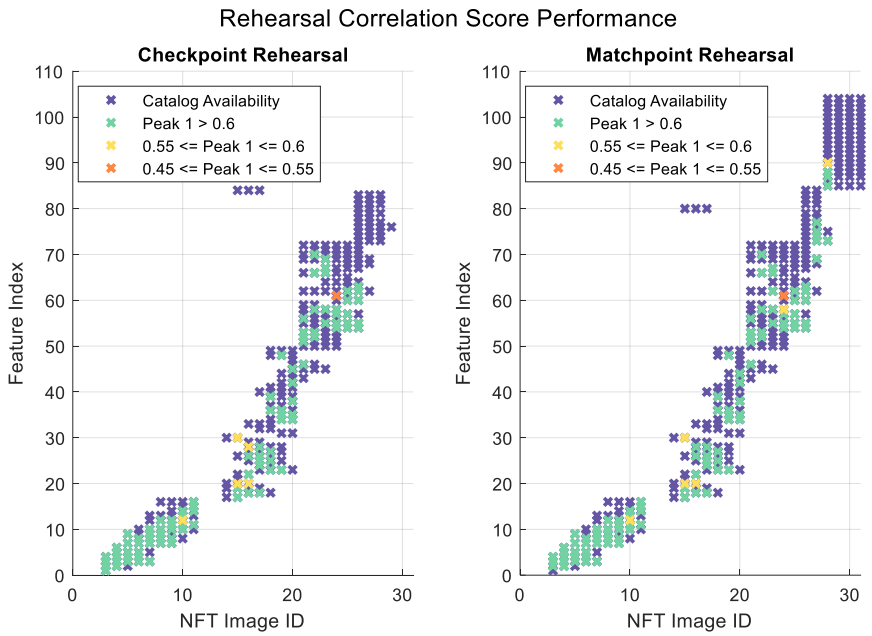


Fig. 12 Feature catalog use and performance during Checkpoint and Matchpoint Rehearsals. In both plots, an 'x' represents a feature that was available to be processed for a single NFT image based on the catalog design. Features used during the rehearsals are colored based on their correlation score performance, where green represents features scoring above 0.6, yellow represents features scoring between 0.55 and 0.6, and orange represents features scoring between 0.45 and 0.55. No features scored less than 0.45.

Figures 13 to 16 show a detailed comparison between test image correlation performance and rehearsal performance. For features that scored lower than expected in rehearsal images, this set of plots was especially useful to allow the NFT team to understand whether the score decrease was due to a change in image conditions. For example, feature 61 scored around 0.5 in the rehearsal images, despite scoring over 0.7 in the test images. Analyzing the image conditions showed that while most test image conditions were similar to rehearsal image conditions, the emission angle during the rehearsals was much larger than tested. The NFT team had concluded from prior testing that large emission angles reduced correlation scores so this behavior was expected and part of why the NFT team pushed the test image correlation scores to be so high. Feature 30 also saw a large decrease in the rehearsals and these plots combined with analysis of the feature renderings suggested this was due to changes in camera angles, especially the camera azimuth angle, compared to the test images. This suggested that this feature may have been biased to the view angles present in OCAMS images used for SPC processing and did not extrapolate as well to other angles. On the other hand, scores increased for some features compared to test images. This increase appeared to be due to the image conditions during the rehearsals being more favorable in either GSD or with smaller camera emission angles for those features. The NFT team studied these plots in this way for every feature, comparing the results with compared feature renderings as needed, to make sure they fully understood feature performance and had confidence features would perform well for TAG.

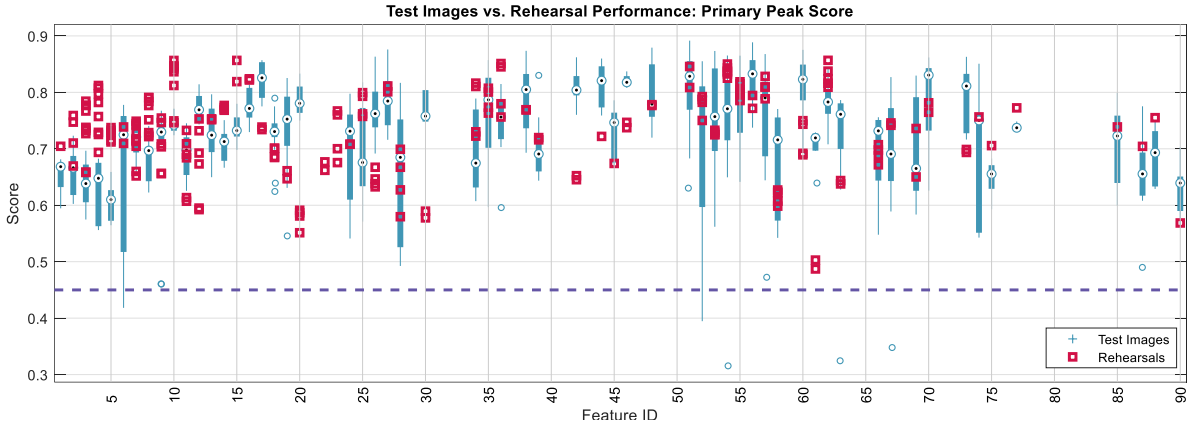


Fig. 13 Feature correlation score performance for features visible in rehearsal images compared to test images. Pink boxes represent correlation scores of the primary peak for both Checkpoint and Matchpoint rehearsal images. Blue data represents the primary peak scores for test images in a box and whisker plot form, where the thick blue sections represent data between the 25th and 75th percentiles, the thin lines extending out represent the most extreme data points considered inliers, and empty blue circles represent outliers. Blue circles with a black dot in the center represent the medians.

Fig. 14 Comparison of GSD between rehearsal images and test images. Plotting styles are the same as defined in Fig. 13.

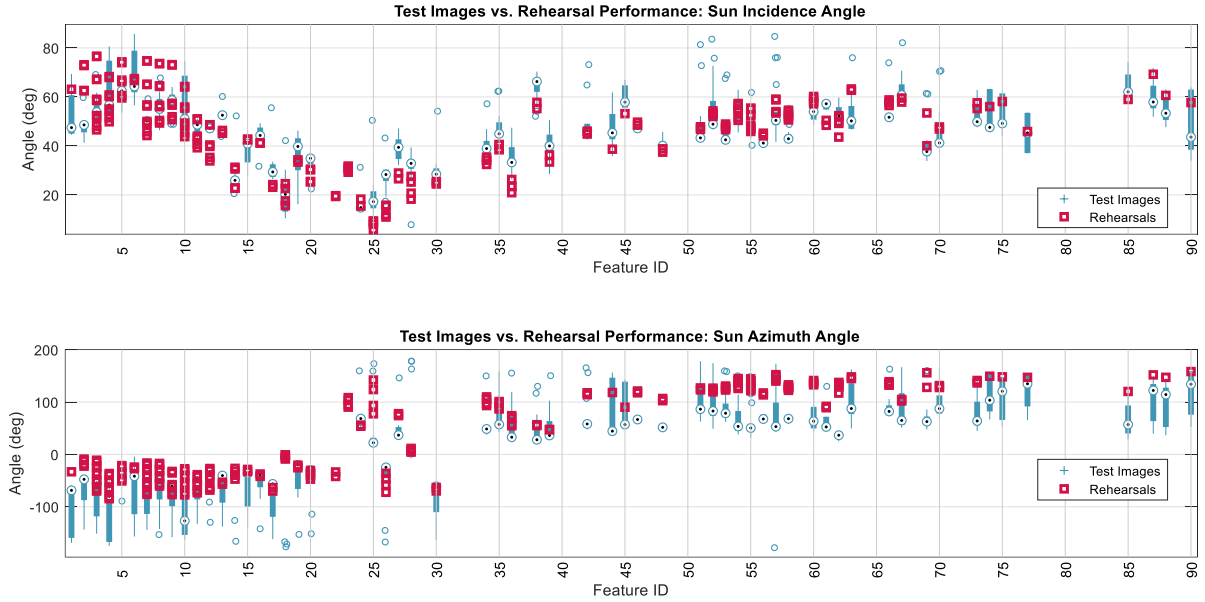


Fig. 15 Comparison of sun angles between rehearsal images and test images. The top plot shows the sun incidence angle while the bottom plot shows the sun azimuth angle. Plotting styles are the same as defined in Fig. 13.

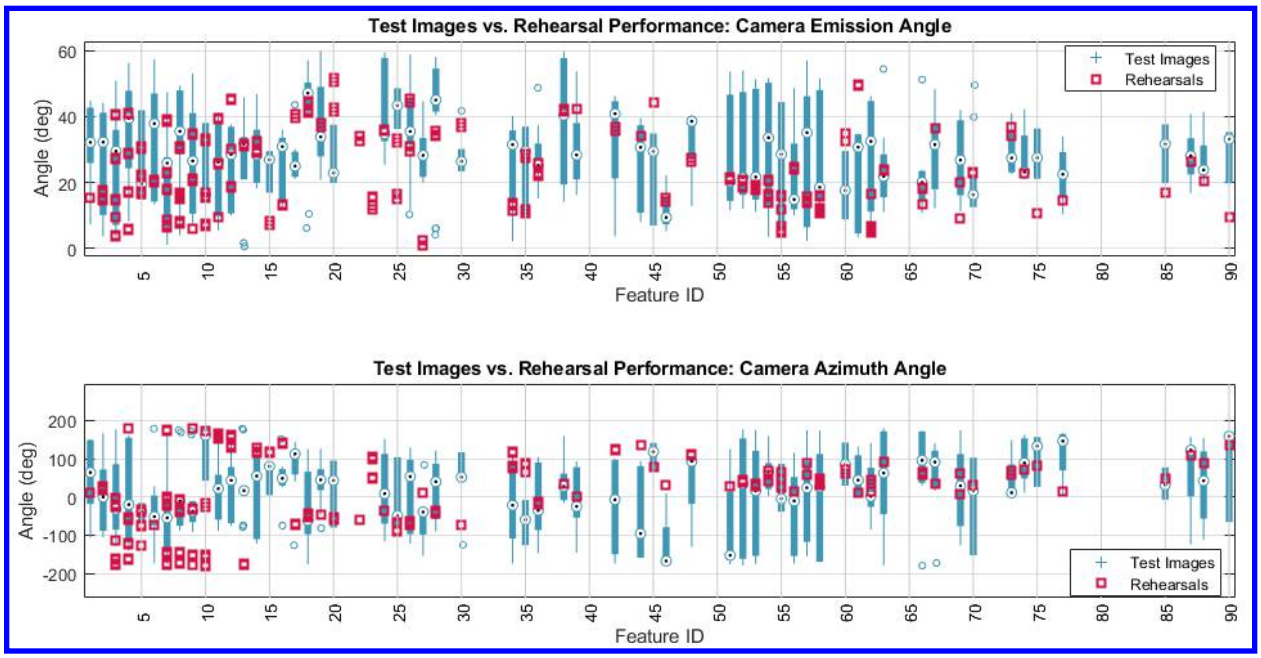


Fig. 16 Comparison of camera angles between rehearsal images and test images. The top plot shows the camera emission angle while the bottom plot shows the camera azimuth angle. Plotting styles are the same as defined in Fig. 13.

Since the rehearsals only tested NFT features through image 28, they did not test the final low altitude NFT images 29-31 which were considered some of the riskiest from a correlation perspective. This risk was primarily due to the small GSD of both the images and features and a lack of test images that supported these GSDs. However, during the Matchpoint Rehearsal, NavCam 2 collected a series of images as the spacecraft backed away from the surface. While not at the GSD of the NFT images 29-31, these images provided one of the best datasets to characterize correlation score changes as a function of only GSD changing, since all other image conditions remained relatively constant

during this spacecraft motion. Analysis from this dataset showed that most features were relatively robust to GSD changes but some saw score decreases as much as 0.05 per 0.6 cm/pix image GSD decrease. Extrapolating this to images 29 to 31 meant that correlation scores could drop from their scores in image 28 by as much as 0.05 to 0.1 through image 31. Applying this score change to the feature scores from image 28 of the Matchpoint Rehearsal and accounting for catalog order and likelihood of which features would be processed suggested that at least 4 features would still pass the 0.45 correlation score threshold per image. This gave the team confidence that features would correlate well through the final NFT image, despite not being able to test features at those image conditions. Feature 90 scored lower for image 28 of the rehearsal with a score of 0.57, so there was still a risk that this feature could drop below 0.45 but this was less of a concern since the team expected to still have 4 passing features. In the end, Feature 90 scored 0.52 for image 30 during TAG, which followed the predicted trend of a score decrease of 0.05 to 0.1 over the final NFT images.

C. Catalog Ordering Process

As mentioned in III.D, the last step to finalize the TAG catalog was to optimize the feature order to ensure that features would correlate robustly and also be spread across the full image field-of-view to ensure good observability. Due to limited development time, there was no time to develop tools to do this autonomously so this process was done manually. The NFT team determined that for the Checkpoint rehearsal, the order in which features were selected for images 1-27 was sufficient, due largely to the fact that these images covered a larger span of terrain and features were already dispersed to limit the amount of processing required for AltWG. After the Checkpoint Rehearsal, some slight changes were made to this ordering based on correlation performance but the catalog stayed mostly the same.

Catalog ordering of features for images 28-31 was more complicated however. This was due to the fact that the majority of features were located in one part of the camera field of view due to the rugged and sloping terrain nearby that prevented NFT from using features in those areas. To order these features, the NFT team first classified features by correlation performance into 4 categories: 1) Features with a high confidence in strong correlation performance, 2) Features with fewer test images but were still expected to correlate well, 3) Features with some lower scoring results but still passed near NFT image conditions, and 4) Features that should be avoided due to poor correlation performance. Along with this correlation classification, the NFT team also considered Monte Carlo image conditions and the expected correlation performance at those conditions. This was made complicated by the fact that while features for these images were separated into one set for images 28-29 and another set for 30-31, many features needed to be used across all 4 images to ensure adequate coverage for all simulated Monte Carlo runs.

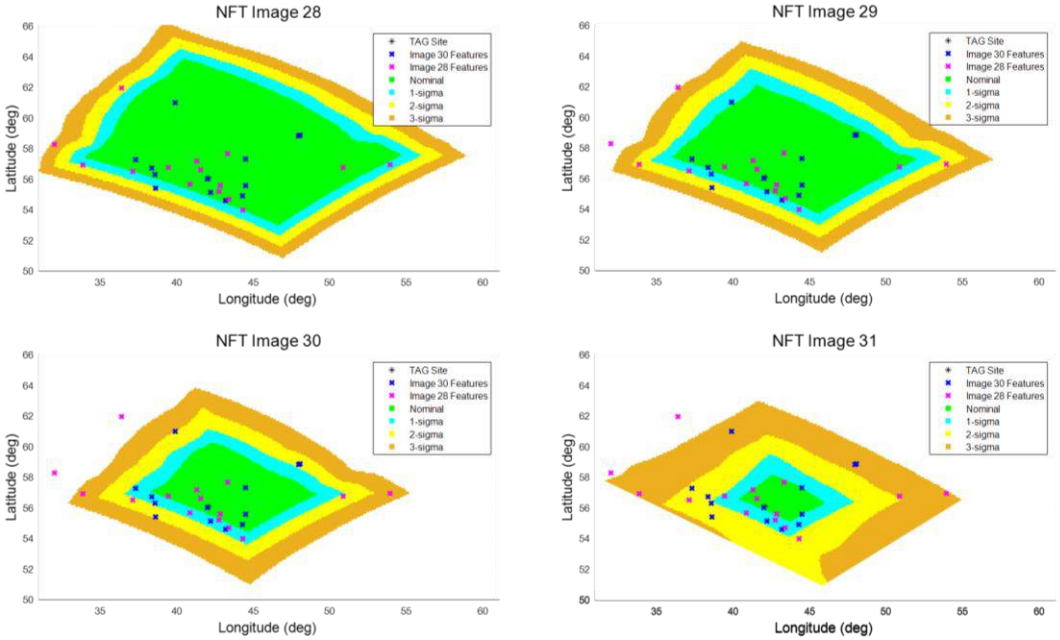


Fig. 17 NFT image footprints and feature locations for images 28-31. Shaded boxes represent the surface coverage area for the nominal (green), 1-sigma (light blue), 2-sigma (yellow), and 3-sigma (brown) trajectory cases. The black asterisk represents the center of the TAG site while the x's represent available feature locations. Pink x's are features designed for images 28-29; blue x's are features designed for images 30-31.

As shown in Fig. 17, the next step was to visually assess where features would be located in the NavCam 2 field of view for images 28-31 for the nominal trajectory and 1-, 2-, and 3-sigma dispersions. The final step was to analyze Monte Carlo data to determine the likelihood that a feature would be visible in each image. The NFT team then leveraged all of this information to manually optimize the catalog order. Figure 18 shows the spread of the features processed in images 28-30 during TAG, demonstrating that the final catalog order resulted in NFT using features that were well-distributed across the images, providing good observability for NFT's state update.

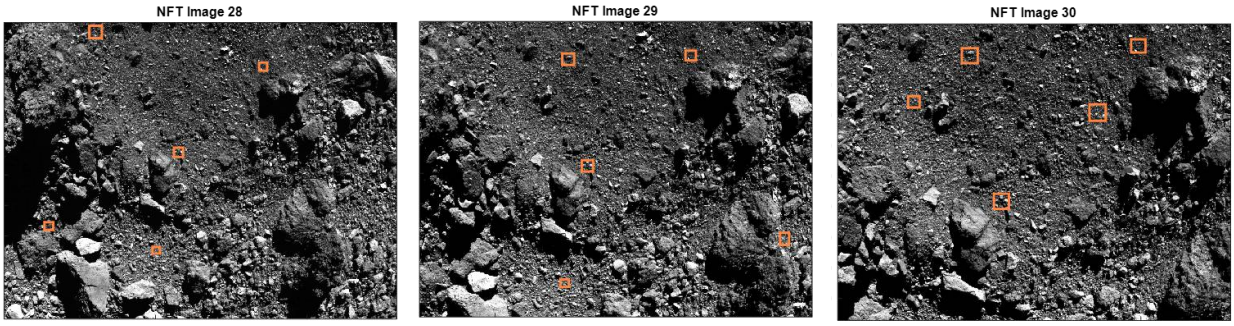


Fig. 18 Features used in NFT images 28-30 during TAG. The top edge of these images corresponds to the lower left edge of the FOVs in the footprints shown in Fig. 17.

VI. Feature Catalog Performance during TAG

NFT features performed extremely well for TAG, with NFT successfully correlating all features processed, even at the lowest altitude images with GSDs smaller than 0.5 cm/pixel. A few features saw lower correlation scores than the OCAMS test results and further analysis has determined this was due to a change in image conditions between the test images and TAG. This was expected and was also the reason why the NFT team added margin to correlation score criteria for OCAMS image testing.

Figure 19 shows the correlation performance of the features used during TAG for both score and match accuracy. While some features have errors approaching 20 pixels, the plot on the right shows that all feature matches were within 0.5 m of the true feature location. Figure 20 compares these primary peak correlation score results to feature performance for the test images. Additional plots are provided in the Appendix that provide data of how the TAG images compared to the test image conditions (similar to Figures 14-16 compare for the rehearsals).

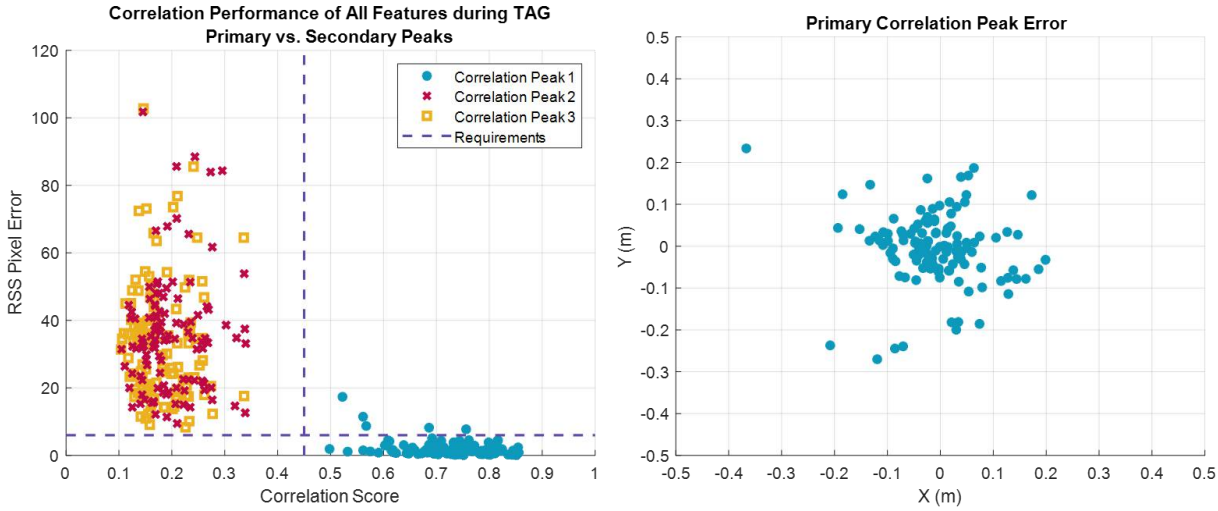


Fig. 19 Correlation performance of features processed during TAG. Correlation peak errors are computed using a reconstructed TAG trajectory to predict feature locations.

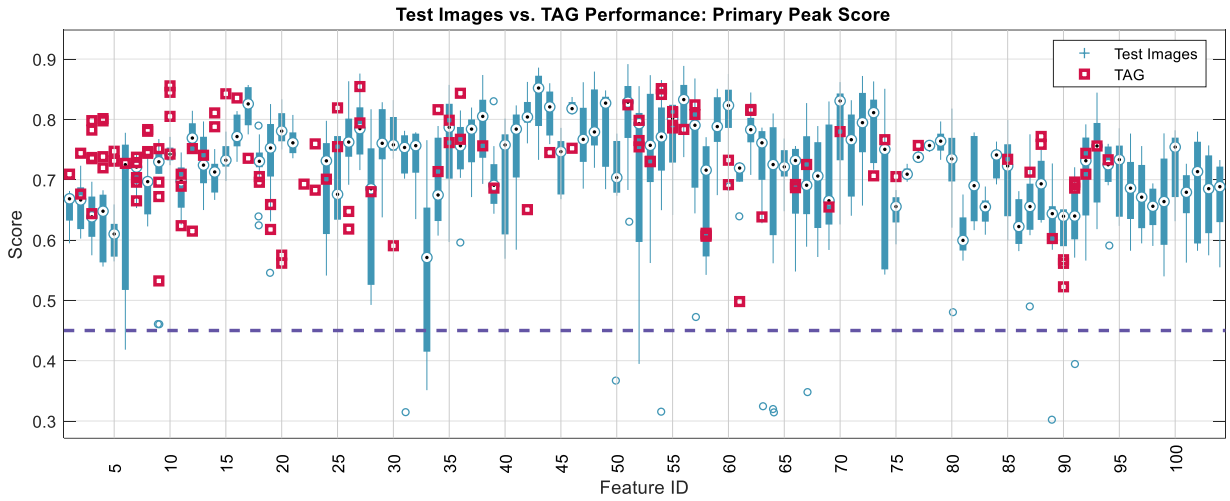


Fig. 20 TAG correlation score performance compared to test images. Plotting styles are the same as defined in Fig. 13, except that here, the pink boxes represent results of features processed during TAG.

Figure 21 provides examples of features that performed well for TAG along with others that saw larger correlation errors or lower scores. On the left, both features score extremely high and with match location errors of only 0.2 cm and 1.8 cm. Looking at the features and image conditions, this appears to be due to small camera emission angles, having TAG images that closely aligned with test images, and using terrain with distinct surface patterns as well as being relatively flat for SPC to be able to model well. On the other hand, the right side shows features that did not perform as well, with scores just above 0.5 and errors on the order of 20 cm. The top feature (feature 9) seems to have performed poorly due both having larger camera emission angles as well as terrain errors that result in incorrectly rendered shadows. Meanwhile the bottom feature (feature 90) struggles largely due to terrain modeling errors that do not correctly represent the albedo or the shapes of the smaller rocks.

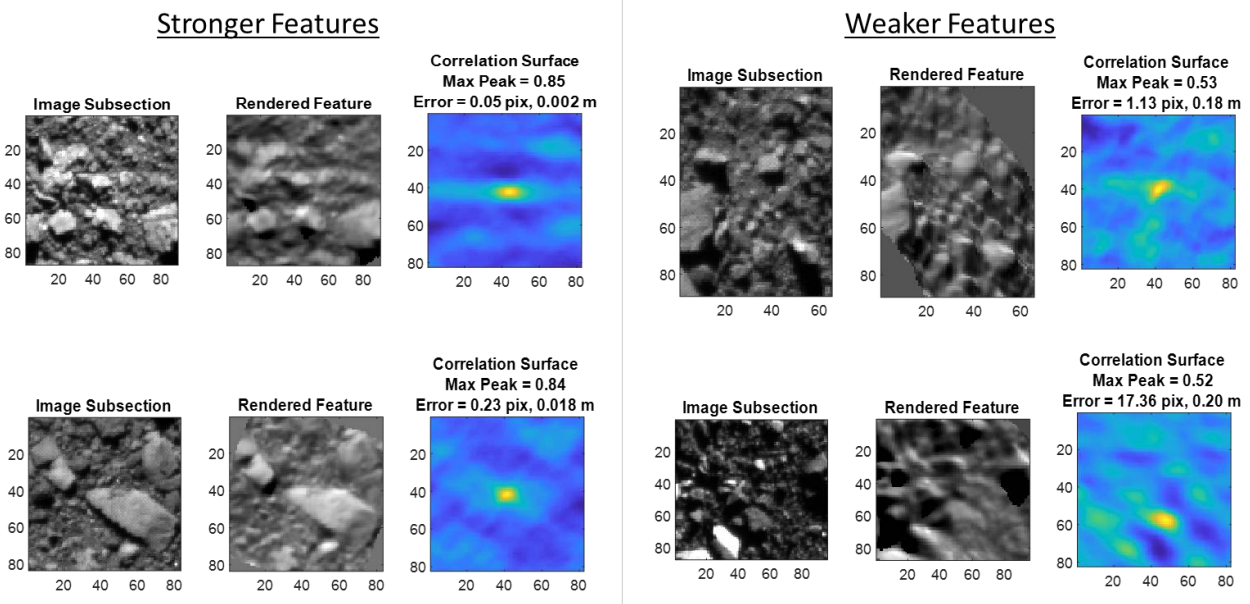


Fig. 21 Example stronger and weaker features during TAG. For each feature, the image subsection is the image patch where the feature matched to, the rendered feature is the output of NFT's onboard rendering software, and the correlation surface is the correlation result. Statistics about correlation match performance for both score and accuracy are provided above the correlation surfaces.

While correlation score performance for these features is dependent mostly on the feature rendering quality, position error could also be due to incorrect feature locations from the OLA DTM alignment process. Analyzing feature 90 further and removing correlation errors that appear to be due to correlation from the size of the peak shows

that the remaining error for all 3 images processed during TAG for this feature is on the order of 12 cm. While the RMS error for this feature after the OLA alignment process was slightly higher compared to other features, AltWG noted that this particular feature was hard to align due to the large GSD difference between the feature and OLA DTM and the lack of any distinct surface topography in the feature to more easily register against the DTM. Given that the OLA DTM was at 4.8 cm/pixel, a 12 cm location error corresponds to only 2.5 pixels of the DTM which is a reasonable amount of error to expect for this particular case. Despite these registration challenges, the majority of feature localization errors were quite low, showing that the AltWG alignment process worked well.

Figure 22 explores this further, showing feature correlation errors during TAG both on a per feature basis (Fig. 22a) and as they compare to the RMS errors computed after the OLA DTM alignment process (Fig. 22b). As might be expected, the plot comparing feature correlation errors to RMS errors shows a potential trend, where smaller RMS errors correlate to smaller correlation errors, further validating that this was a useful metric for AltWG to use to assess the alignment process.

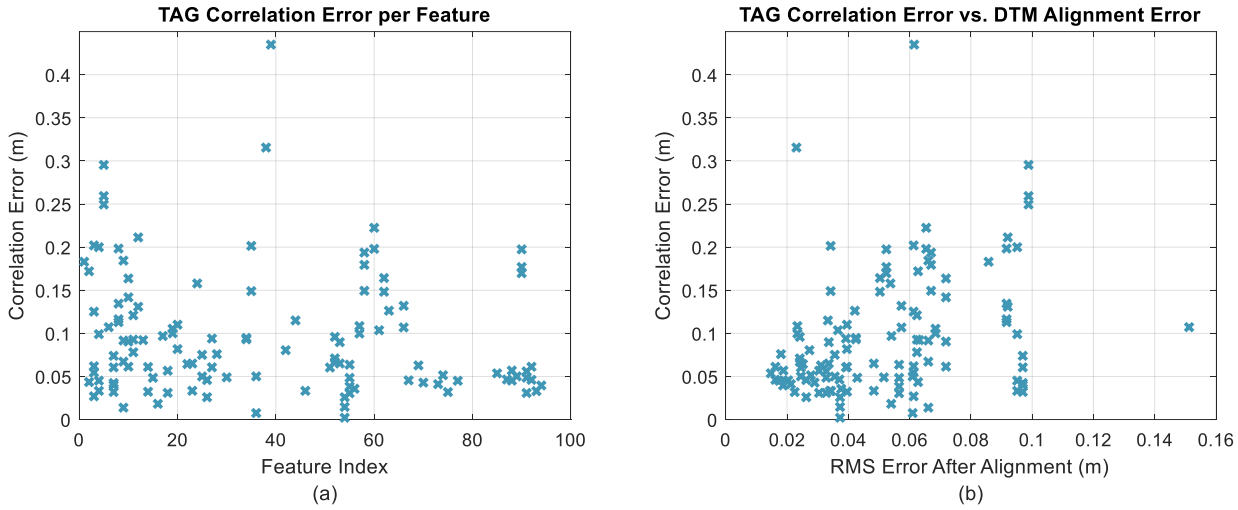


Fig. 22 Feature correlation error during TAG. Correlation error is the error between the detected feature location and the feature location and could be due to either correlation errors from inaccurately rendered features or errors in defined feature locations. (a) Correlation error per feature. (b) Correlation error compared to feature RMS error after alignment to OLA DTM.

VII. Conclusions

The OSIRIS-REx sample collection event was successful in large part due to the extensive NFT feature catalog development process performed by both the NFT and AltWG teams. With only room to process five features per image and therefore little room for outlier feature matches, NFT successfully correlated all features, correctly matching every feature without a single false match. These feature matches enabled NFT to guide the spacecraft to within 1 m of the targeted sample site and produce a final estimate of the contact location that was only 3.5 cm from where the spacecraft ultimately made contact.

The results presented in this paper show that the NFT feature development and validation strategies were effective for designing a robust feature catalog that enabled NFT to successfully navigate for the sample collection. Selecting features that could be modeled well by SPC entailed selecting terrain that was relatively smooth, without large height variations, included unique rock and/or albedo patterns, and had sufficient orbital image coverage to provide a wide range of image conditions. While testing was challenging due to the fact that available OCAMS test images did not match expected TAG conditions simultaneously for all image conditions, the team developed a feature evaluation strategy that analyzed all aspects of feature performance and renderings. Despite being manually intensive, this process improved the team's understanding of SPC performance for Bennu's rugged surface, increased confidence in feature performance for TAG, and helped the team predict score performance for features at TAG conditions. Ultimately, one of the most effective feature evaluation strategies was to add margin to the correlation score threshold to ensure that the feature would still score above NFT's onboard threshold even if correlation performance was worse under TAG conditions. Lastly, both the Checkpoint and Matchpoint Rehearsals were extremely valuable exercises to validate feature performance, predict performance for the final NFT images, and optimize the catalog order.

Appendix

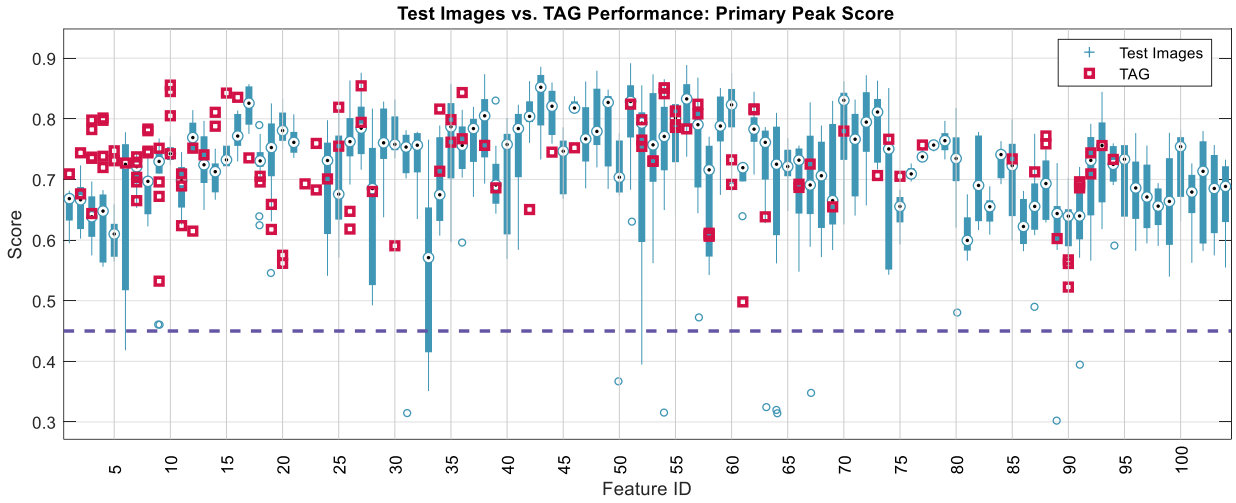


Fig. 23 Feature correlation score performance for features used during TAG compared to test images. Pink boxes represent primary peak correlation scores during TAG. Blue data represents the primary peak scores for test images in a box and whisker plot form, where the thick blue sections represent data between the 25th and 75th percentiles, the thin lines extending out represent the most extreme data points considered inliers, and empty blue circles represent outliers. Blue circles with a black dot in the center represent the medians.

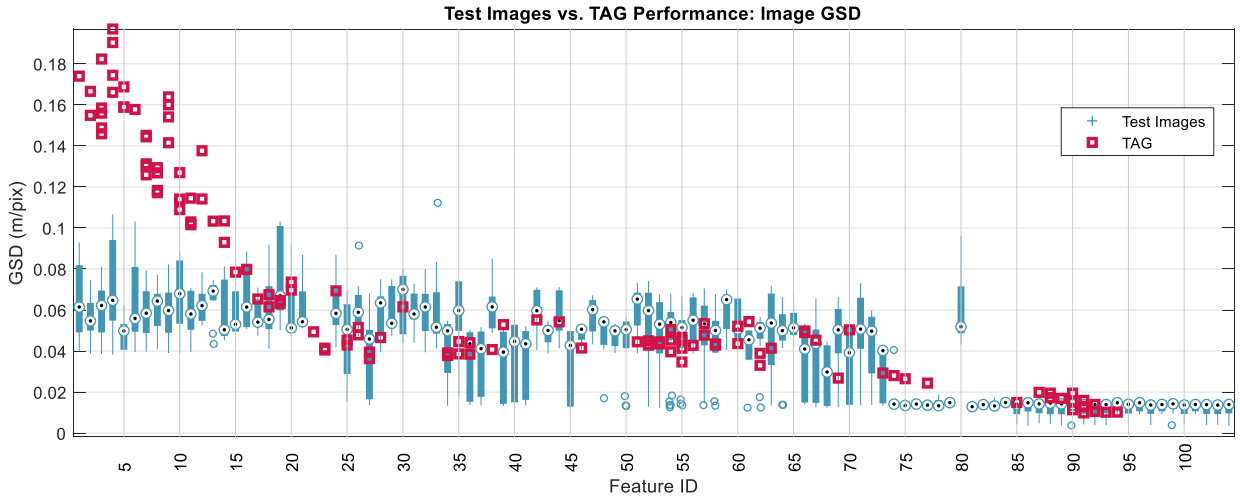


Fig. 24 Comparison of GSD between rehearsal images and test images. Plotting styles are the same as defined in Fig. 23.

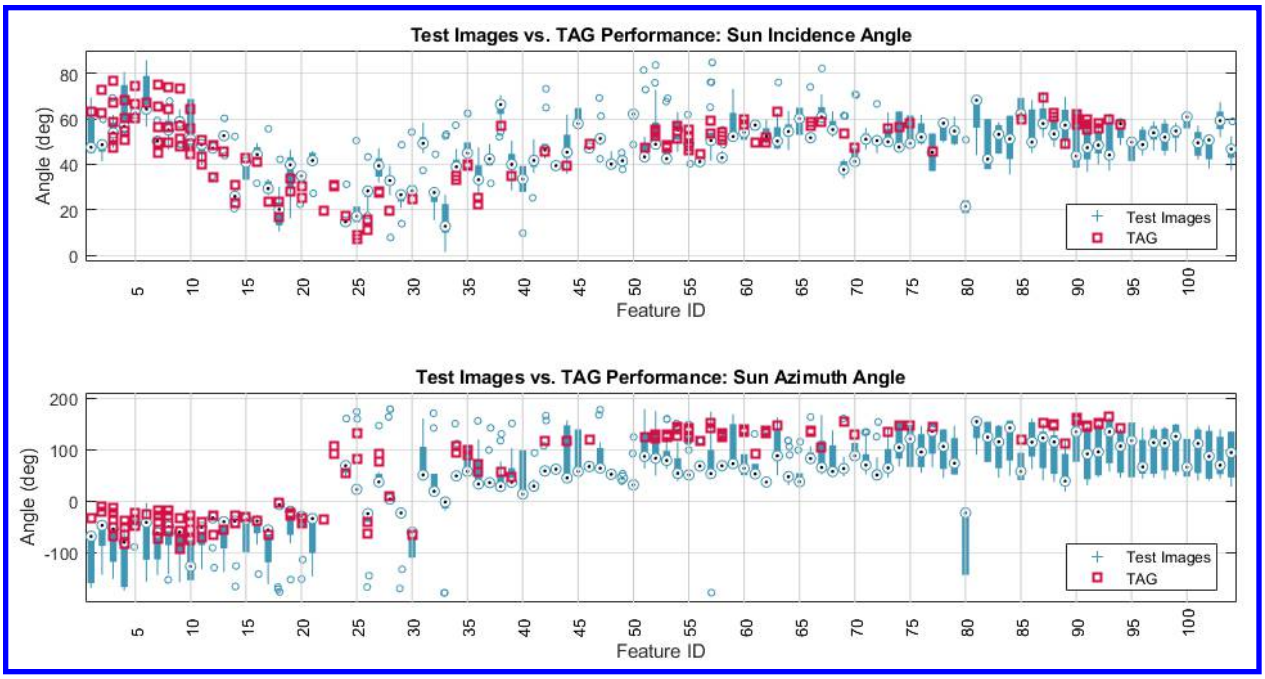


Fig. 25 Comparison of sun angles between rehearsal images and test images. The top plot shows the sun incidence angle while the bottom plot shows the sun azimuth angle. Plotting styles are the same as defined in Fig. 23.

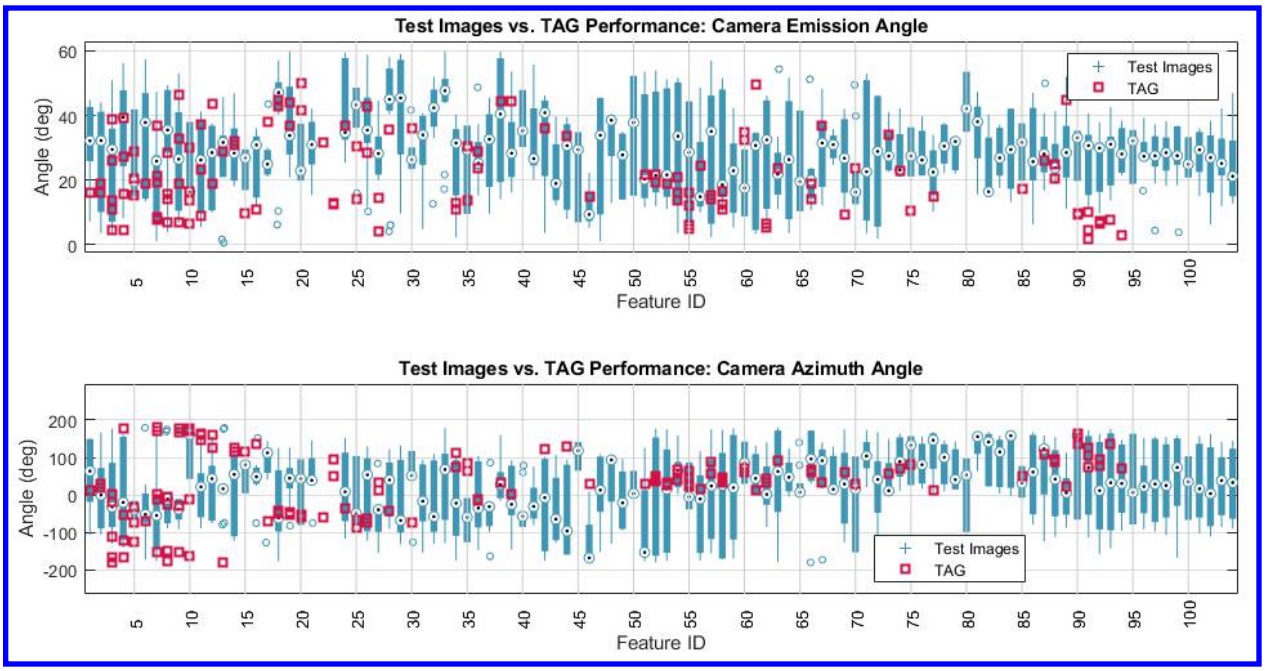


Fig. 26 Comparison of camera angles between rehearsal images and test images. The top plot shows the camera emission angle while the bottom plot shows the camera azimuth angle. Plotting styles are the same as defined in Fig. 23.

Acknowledgements

This work was authored by employees of Draper under Contract NNJ16HK08B with the National Aeronautics and Space Administration and is based upon work supported by NASA under Contracts NNM10AA11C and NNG12FD66C issued through the New Frontiers Program. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, worldwide license to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, or allow others to do so, for United States Government purposes. All other rights are reserved by the copyright owner.

Funding for the Canadian team members and for OLA was provided by the Canadian Space Agency. We are grateful to the entire OSIRIS-REx Team for making the encounter with Bennu possible. OLA (Daly et al. 2019), OCAMS (Rizk et al. 2019), and NavCam/TAGCAMS (Bos et al. 2019) data are available via the Planetary Data System (https://sbn.psi.edu/pds/resource/orex/).

References

[1] Lauretta, D. S. and the OSIRIS-REx TAG Team, "The OSIRIS-REx Touch-and-Go Sample Acquisition Event and Implications for the Nature of the Returned Sample," *Lunar and Planetary Science Conference*, 2021.

[2] Bierhaus, E.B., Clark, B.C., Harris, J.W. et al, "The OSIRIS-REx Spacecraft and the Touch-and-Go Sample Acquisition Mechanism (TAGSAM)," *Space Sci Rev* 214, 107 (2018). doi: 10.1007/s11214-018-0521-6

[3] Olds, R. D., Miller, C. J., Norman, C. D., Mario, C. E., Berry, K., Palmer, E., Barnouin, O. S., Weirich, J. R., Lauretta, D. S. "Use of Shape Models for Natural Feature Tracking at Asteroid Bennu", *Planetary Science Journal* (in preparation).

[4] Norman, C. D., Miller, C., Olds, R., Mario, C., Palmer, E., Barnouin, O. S., Weirich, J. S., Bos, B. J., Rizk, B., Lauretta, D. S., "Natural Feature Tracking Performance During the OSIRIS-REx Touch-And-Go Event," *Planetary Science Journal* (in preparation).

[5] Lauretta, D.S. et al., 2017. OSIRIS-REx: Sample Return from Asteroid (101955) Bennu. Space Science Reviews 212, 925–984.

[6] Lauretta, D. S. et al., 2021. OSIRIS-REx at Bennu: Overcoming challenges to collect a sample of the early Solar System. In Sample Return Missions, ed. Longobardo, A. (Elsevier), pp. 163–194.

[7] Bos, B. J., Ravine, M. A., Caplinger, M., et al., "Touch And Go Camera System (TAGCAMS) for the OSIRIS-REx Asteroid Sample Return Mission," *Space Sci Rev* 214, 37, 2018. doi: 10.1007/s11214-017-0465-2

[8] Rizk, B., Drouet d'Aubigny, C., Golish, D. et al., "OCAMS: The OSIRIS-REx Camera Suite," *Space Sci Rev* 214, 26 (2018). doi: 10.1007/s11214-017-0460-7

[9] Daly, M.G., Barnouin, O.S., Dickinson, C., Seabrook, J., Johnson, C.L., Cunningham, G., Haltigin, T., Gaudreau, D., Brunet, C., Aslam, I., Taylor, A., Bierhaus, E.B., Boynton, W., Nolan, M., Lauretta, D.S., 2017. The OSIRIS-REx Laser Altimeter (OLA) Investigation and Instrument. Space Science Reviews 212, 899–924. https://doi.org/10.1007/s11214-017-0375-3

[10] Jackman, C.D., et al., 2017. Optical navigation concept of operations for the OSIRIS-REx mission Advances in the Astronautical Sciences. Spaceflight Mechanics 2017. Univelt, AAS 17-489 vol. 160.

[11] Barnouin, O.S., Daly, M.G., Palmer, E.E., Johnson, C.L., Gaskell, R.W., Asad, M.A., Bierhaus, E.B., Craft, K.L., Ernst, C.M., Espiritu, R.C., Nair, H., Neumann, G.A., Nguyen, L., Nolan, M.C., Mazarico, E., Perry, M.E., Philpott, L.C., Roberts, J.H., Steele, R.J., Seabrook, J., Susorney, H.C.M., Weirich, J.R., Lauretta, D.S., 2020. Digital terrain mapping by the OSIRIS-REx mission. Planetary and Space Science 180, 104764. https://doi.org/10.1016/j.pss.2019.104764

[12] Johnson, A., Aaron, S., Chang, J., Cheng, Y., Montgomery, J., Mohan, S., Schroeder, S., Tweddle, B., Trawny, N., Zheng, J., "The Lander Vision System for Mars 2020 Entry Descent and Landing *AAS Guidance, Navigation, and Control Conference*, AAS-17-038, February 2017.

[13] Dever, C., Dyer, T., Hamilton, L., Lommel, P., Mohiuddin, S., Reiter, A., Singh, N., Truax, R., Wholey, L., "Guided-Airdrop Vision-Based Navigation," AIAA 2017-3723. 24th AIAA Aerodynamic Decelerator Systems Technology Conference. June 2017.

[14] Owens, C., Macdonald, K., Hardy, J., Lindsay, R., Redfield, M., Bloom, M., Bailey, E., Cheng, Y., Clouse, D., Villalpando, C.Y., Hambardzumyan, A., Johnson, A.E, Horchler, A.D. "Development of a Signature-based Terrain Relative Navigation System for Precision Landing," AIAA 2021-0376. AIAA Scitech 2021 Forum. January 2021.

[15] Singh, L., Lim, S.. "On Lunar On-Orbit Vision-Based Navigation: Terrain Mapping, Feature Tracking Driven EKF," AIAA 2008-6834. AIAA Guidance, Navigation and Control Conference and Exhibit. August 2008.

[16] Downes, L.M., Steiner, T.J., How, J.P. (2021). Neural Network Approach to Crater Detection for Lunar Terrain Relative Navigation. Journal of Aerospace Information Systems. 18. 1-13. doi: 10.2514/1.I010884.

[17] Christian, J.A., Derksen, H., Watkins, R. Lunar Crater Identification in Digital Images. J Astronaut Sci (2021). doi: 10.1007/s40295-021-00287-8
[18] Olds, R., May, A., Mario, C., Hamilton, R., Debrunner, C., Anderson, K., "The Application of Optical Based Feature Tracking to OSIRIS-REx Asteroid Sample Collection," *AAS Guidance, Navigation, and Control Conference,* AAS 15-114, 2015.
[19] Everett, D., Mink, R., Linn, T., Wood, J., "Designing to sample the unknown: Lessons from OSIRIS-REx project systems engineering," 2017 IEEE Aerospace Conference, 2017, pp. 1-19. doi: 10.1109/AERO.2017.7943586.
[20] Mario, C., Debrunner, C., "Robustness and Performance Impacts of Optical-Based Feature Tracking to OSIRIS-REx", *AAS Guidance, Navigation, and Control Conference,* AAS 16-087, 2016.
[21] Mario, C., Norman, C., Palmer, E., Weirich, J., Miller, C., Lorenz, D. A., Olds, R., Lauretta, D. S., "Image Correlation Performance Prediction for Autonomous Navigation of OSIRIS-REx Asteroid Sample Collection", *AAS Guidance, Navigation, and Control Conference,* AAS 20-087, 2020.
[22] Gaskell, R.W., Barnouin, O. S., Scheeres, D. J., et al., "Characterizing and navigating small bodies with imaging data," *Meteoritics and Planetary Science*, 43(6), pp.1049–1061, 2008. doi: 10.1111/j.1945-5100.2008.tb00692.x
[23] Palmer, E. E. et al, "How to SPC?," *Planetary Science Journal* (in preparation).
[24] Daly, M.G., Barnouin, O.S., Dickinson, C. et al., "The OSIRIS-REx Laser Altimeter (OLA) Investigation and Instrument," *Space Sci Rev* 212, 899–924 (2017). doi: 10.1007/s11214-017-0375-3
[25] Seabrook, J.A., Daly, M.G, Barnouin, O.S., et al., "Global shape modeling using the OSIRIS-REx scanning Laser Altimeter," *Planetary and Space Science*177, 104688, 2019. doi: 10.1016/j.pss.2019.07.003
[26] Mario, C. E., Miller, C. J., Norman, C. D., et al. "Shape Model Ground Testing to Prepare for OSIRIS-REx Autonomous Vision Navigation using Natural Feature Tracking," *Planetary Science Journal* (in preparation).
[27] Lauretta, D.S., DellaGiustina, D.N., et al., 2019. The unexpected surface of asteroid (101955) Bennu. Nature 568, 55–60.
[28] DellaGiustina, D. N., Burke, K. N., Walsh, K. J., et al., "Variations in color and reflectance on the surface of asteroid (101955) Bennu," *Science* 370, .eabc3660, 2020. doi: 10.1126/science.abc3660
[29] Miller, C., Olds, R., Norman, C., Gonzales, S., Mario, C., Lauretta, D.S., "On-Orbit Evaluation of Natural Feature Tracking for OSIRIS-REx Sample Collection", *AAS Guidance, Navigation, and Control Conference,* AAS 20-154, 2020.
[30] Daly, M.G., Barnouin, O.S., Seabrook, J.A., Roberts, J., Dickinson, C., Walsh, K.J., Jawin, E.R., Palmer, E.E., Gaskell, R., Weirich, J., Haltigin, T., Gaudreau, D., Brunet, C., Cunningham, G., Michel, P., Zhang, Y., Ballouz, R.L., Neumann, G., Perry, M.E., Philpott, L., Al Asad, M.M., Johnson, C.L., Adam, C.D., Leonard, J.M., Geeraert, J.L., Getzandanner, K., Nolan, M.C., Daly, R.T., Bierhaus, E.B., Mazarico, E., Rozitis, B., Ryan, A.J., Dellaguistina, D.N., Rizk, B., Susorney, H.C.M., Enos, H.L., Lauretta, D.S., 2020. Hemispherical differences in the shape and topography of asteroid (101955) Bennu. Science Advances 6. https://doi.org/10.1126/sciadv.abd3649