ESA ASTRA 2023

VISION-BASED LOCALIZATION FOR THE MSR SAMPLE TRANSFER ARM

Marcos Avilés[(1)], David Savary[(1)], Augusto Gómez[(1)], Marco Mammarella[(1)]
Andrea Rusconi[(2)], Francesco Villa[(2)], Guido Sangiovanni[(2)], Davide Nicolis[(3)]

(1) GMV Aerospace & Defence SAU, Spain
(2) Leonardo SpA, Italy
(3) ESA/ESTEC, Netherlands

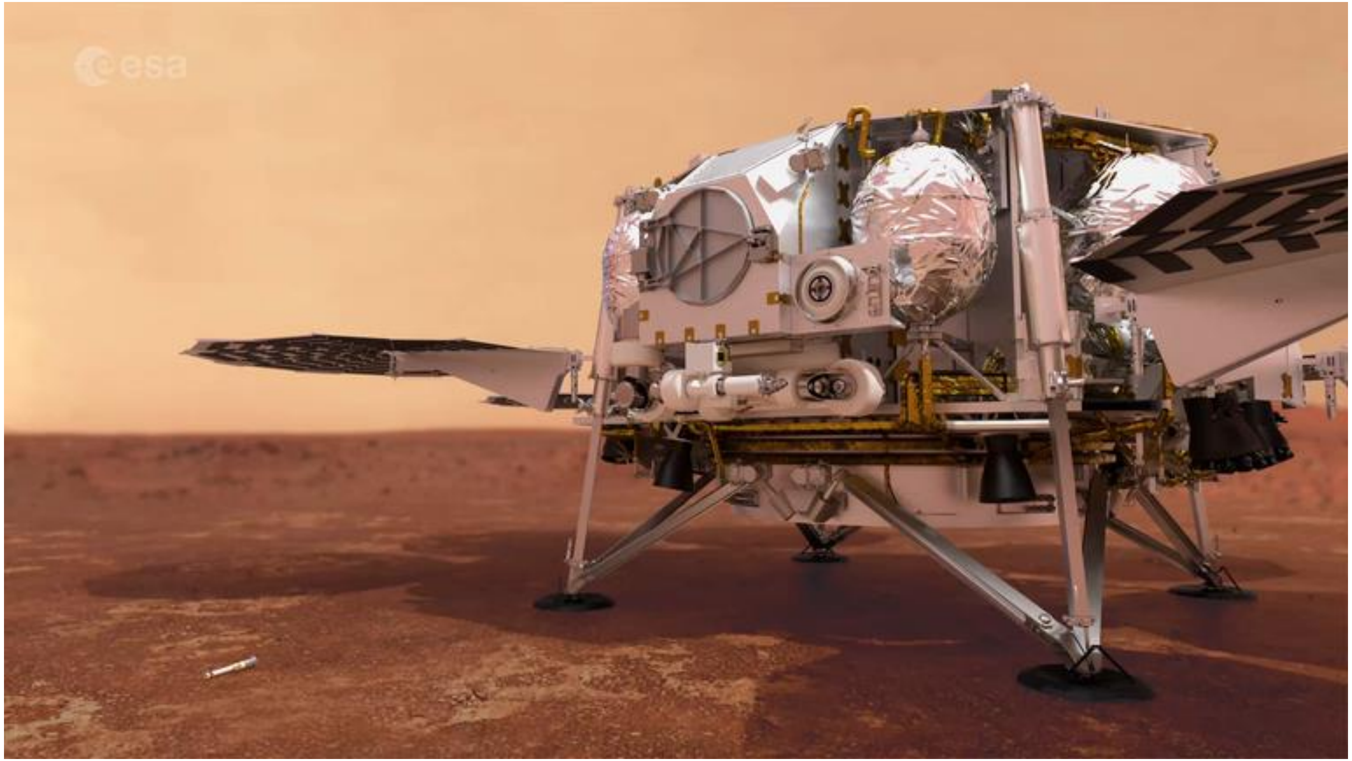
INNOVATING SOLUTIONS

Introduction

- The Mars Sample Return - Sample Transfer Arm is one of the ESA contributions to the MSR Campaign, and is part of the Sample Retrieval Lander
- The STA will:
  - Transfer the sample tubes into the Orbiting Sample container (OS) in the Mars Ascent System (MAS), from:
    - The Perseverance rover, or
    - Mars terrain (dropped by the Sample Recovery Helicopters)
  - Close and secure the OS lid after the completion of the tube transfer operations
- These operations need to be performed within a limited time → high level of autonomy needed
- The vision algorithms for accurate localization of the target elements to be manipulated are key elements to this autonomy

STA Mission Operations

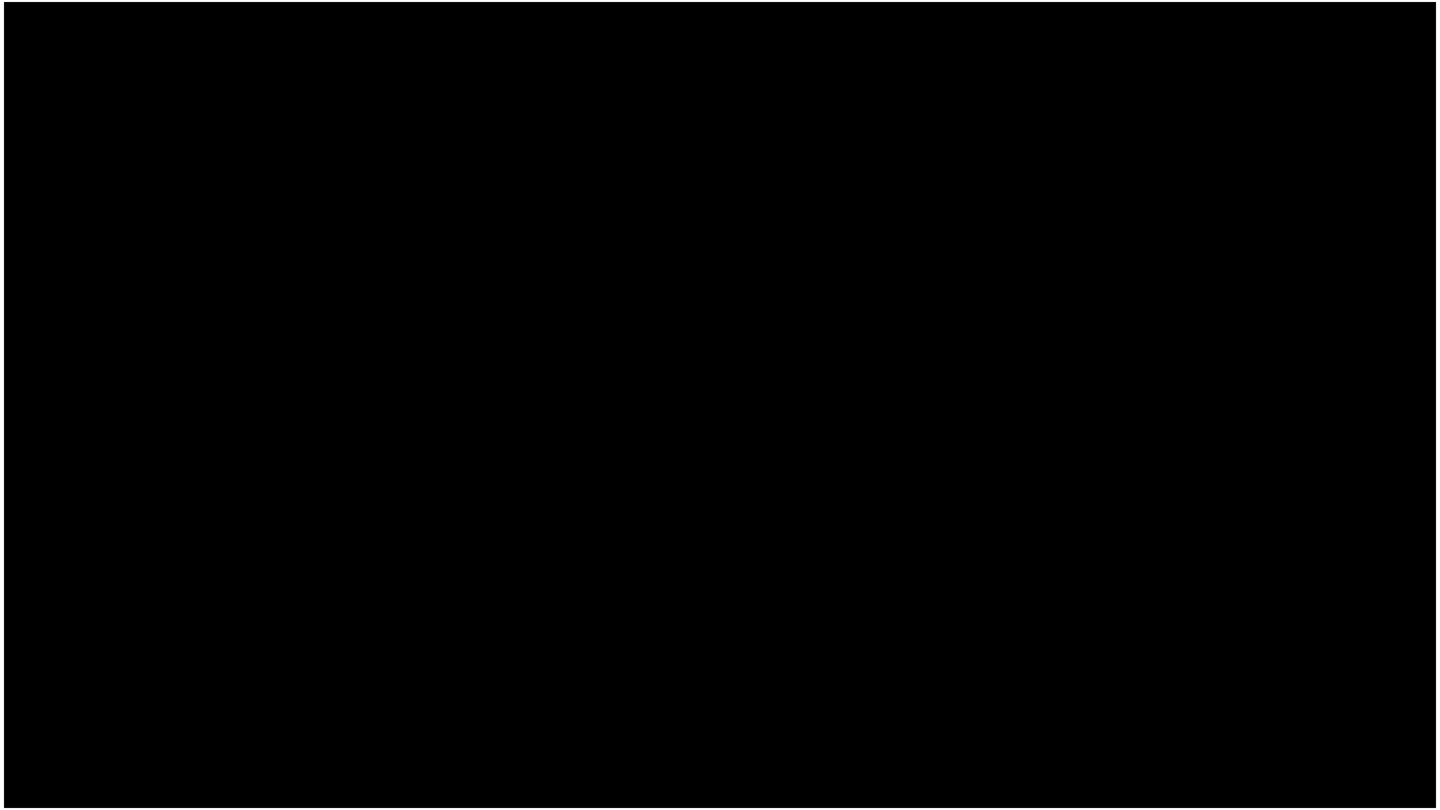
Credit: ESA



gmv

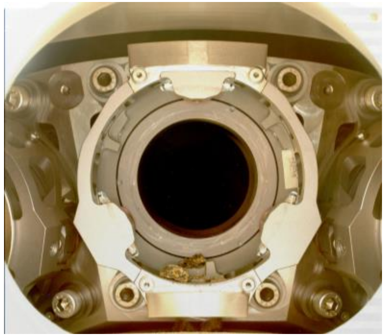
STA Mission Operations

Credit: NASA

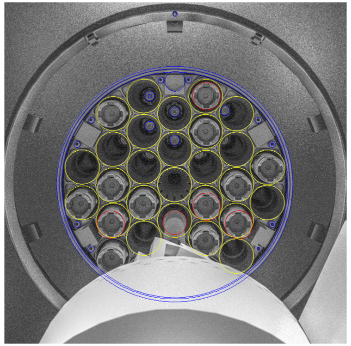


Vision Algorithms in the STA

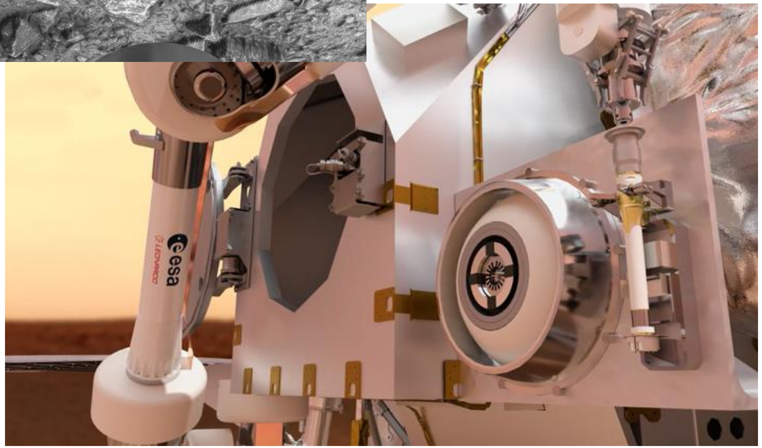
- This work focuses on the design, development and initial testing of the vision algorithms for:
  - Localization of the **Perseverance Bit Carousel**, from where the STA collects the sample tubes
  - Localization of the **OS container**, where the STA inserts the sample tubes
  - Localization of **sample tubes** on the terrain left by the Sample Recovery Helicopters
  - Localization of the **OS Lid**, to be placed onto the OS
  - Localization of the **Workbench**, where the tubes are placed to switch the STA end effector grip type



Credit: NASA



Credit: ESA



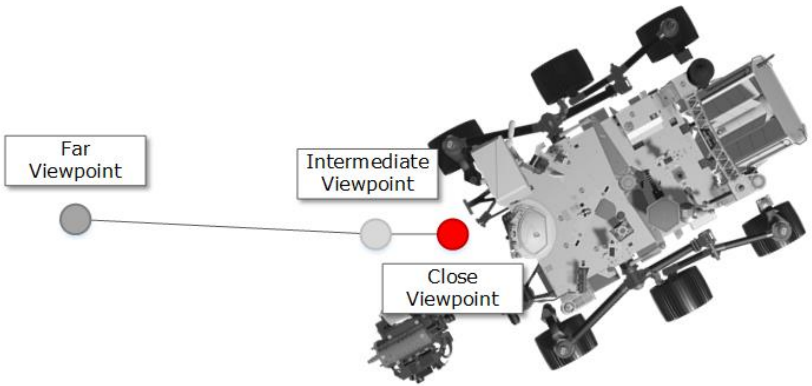
Perseverance & OS Localization Con-Ops

- Due to the uncertainties in the positions of the different elements to be operated, vision-based localization is performed incrementally, at three different arm camera viewpoints:

  - **Far Viewpoint**. Based on a (ground-based) teach point location, to compute a more accurate estimation that allows a safer approach to the target

  - **Intermediate Viewpoint**. Based on the previous estimate the arm can safely get closer to the target within the limits of the accuracy of the previously computed pose

  - **Close Viewpoint**. Based on the already accurate estimate obtained at the medium point, the arm moves as close as possible to the target to compute the final and most accurate estimate



Perseverance & OS Localization Pipeline

- Both scenarios follow a similar pipeline composed of three steps:

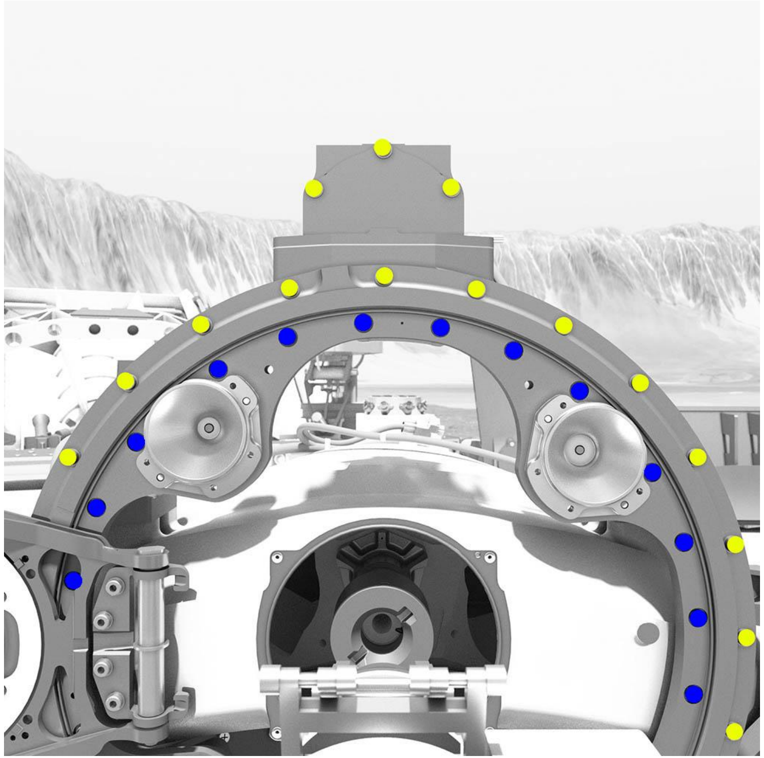
  - **Feature Detection**: Relevant features which could be identified both in the image and in the CAD model are detected

  - **Feature Matching**: Previously detected features are matched against a reference model (obtained from the CAD) to associate their 2D coordinates in the image with their 3D position in the model

  - **Pose Estimation**: Based on the 2D-3D correspondences the 6DOF position of the camera is recovered by solving the Perspective-N-Point

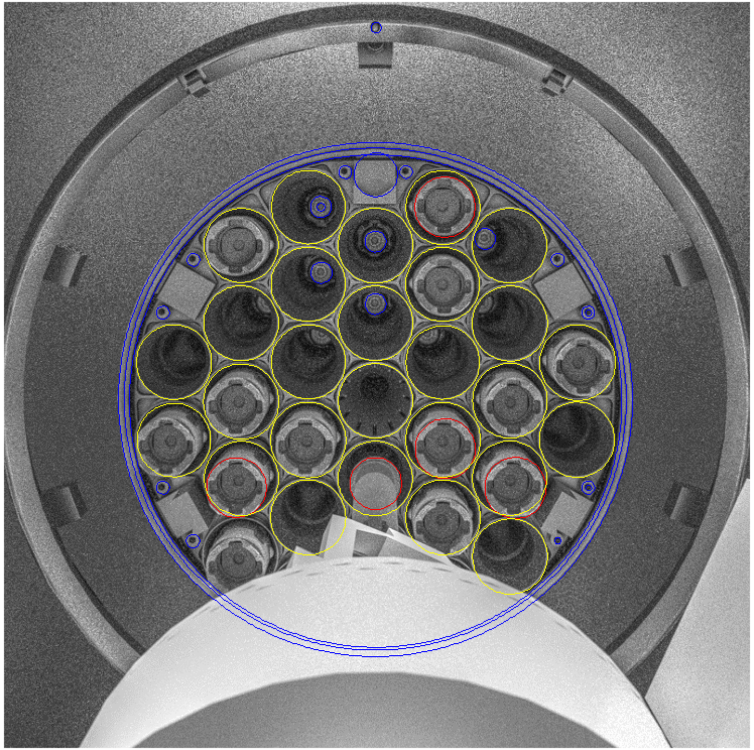
Perseverance Localization

- Fasteners in the Bit Carousel are used as features:
  - Smaller size implies lower error (error for larger circles could be small relative to their size, but large as an absolute error in px)
  - They can be detected with reliability
- Both inner ring (blue) and outer ring (yellow) fasteners are used for robust matching to recover from large input errors.
  - Assumes a coarse knowledge of the inner ring orientation
- Only outer ring fasteners, which are fixed, are used for pose estimation



OS Localization

- For the OS, the circular slots are used as features
- Multiple circles might be detected for the same slot due to the presence of RSTA, shadows,...
  - Filtering is applied to remove circles with sizes incompatible with the viewing distance
  - A grouping stage is performed to merge clusters of circles



Sample Tube Localization

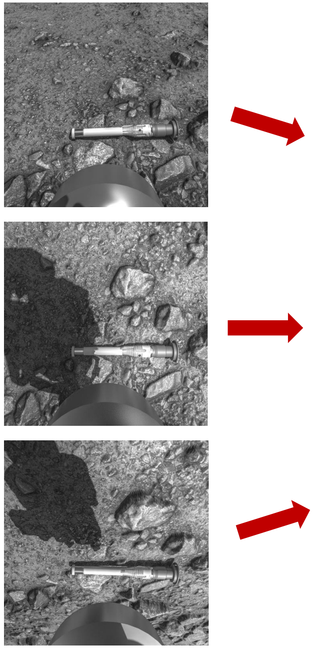
- Computing the Sample Tube pose from a single image is a challenging problem due to the lack of observability, particularly for the inclination over the terrain and the Z distance (depth)

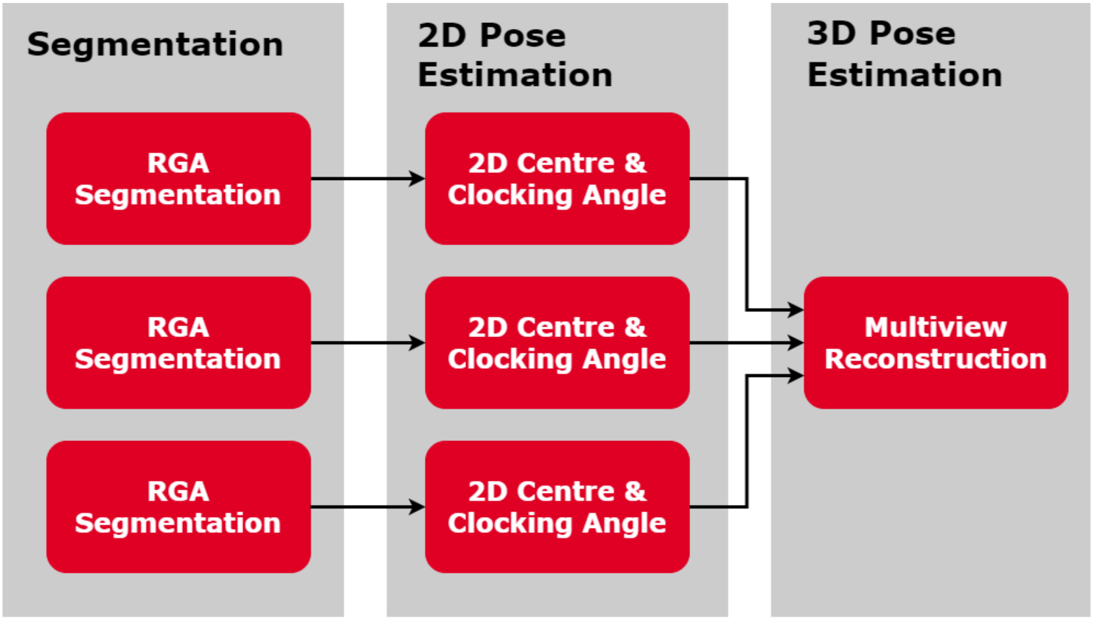
- No clear distinct features can be determined which could be detected in the image and matched against corresponding 3D features in the model

- A multi view capture from 3 different camera orientations is proposed to provide a feasible solution for the 5DOF pose estimation of the tube

Vision Algorithms – Sample Tube Localization

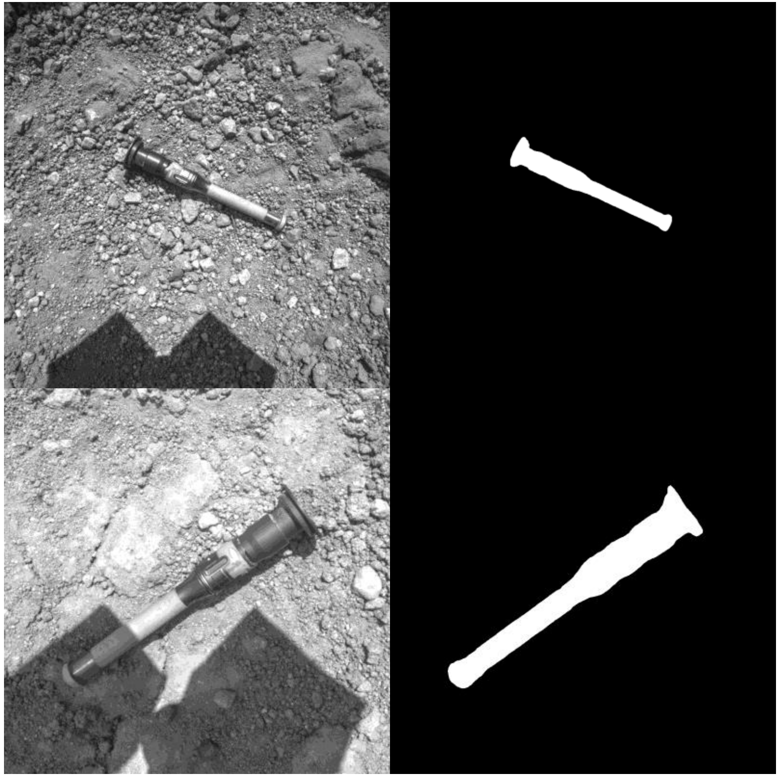
- Algorithm pipeline





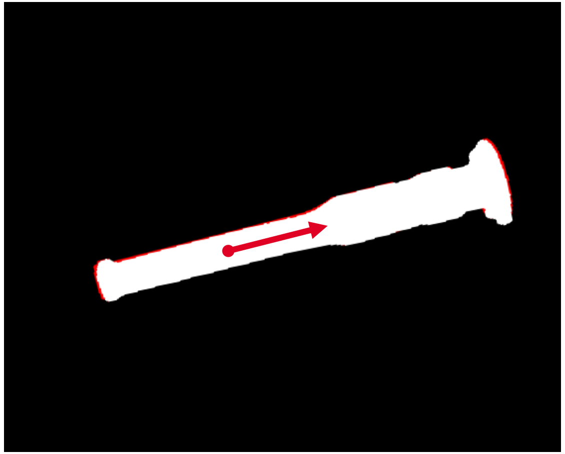
Sample Tube Segmentation

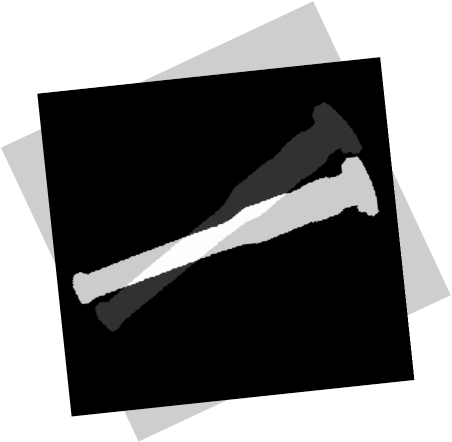
- Segmentation performed using U-Net CNN
  - Encoder section replaced with a pretrained version based on ImageNet
- Model trained with real images of the tubes on sandboxes and computed generated scenes
- Ground truth mask generation of real images performed with a custom semi-automatic tool based on Segment Anything
- Data augmentation pipeline containing random flips, rotations, and perspective transformations
- Good performances with only hundreds of real images and a relatively compact model (5.5 million parameters)



Sample Tube Localization (2D)

- 2D centre and angle estimated through the registration (via ICP) of a reference mask and the result of the segmentation step.
- The reference mask of the Sample Tube is computed based on an initial guess of the distance from the camera.
- The ICP is iterated with slightly scaled versions to get the best fit with the model.



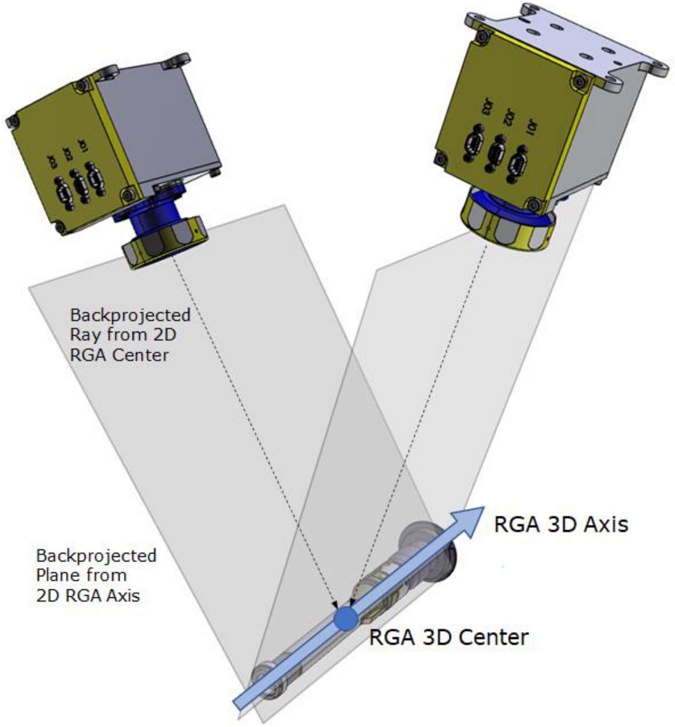


Sample Tube Localization (3D)

- Inputs:
  - Sample Tube 2D position and angle on the images
  - Telemetry from the robotic arm at the different captures
  - Camera intrinsic calibration

- Multi-view reconstruction and estimation of 5DOF tube pose:
  - The tube centre point is computed by intersecting the vectors from each of the camera positions to the 2D centre points
  - The 3D axis of the tube is computed by intersecting the planes that contain the main axis and the camera positions

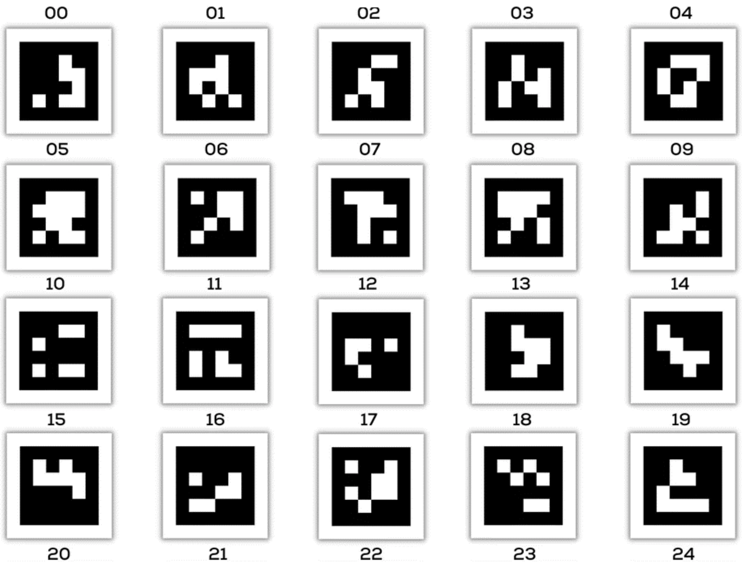
Note: Due to the erroneous nature of the measurements, no exact intersection really exists. The returned value is the one that minimizes the distances to rays/planes



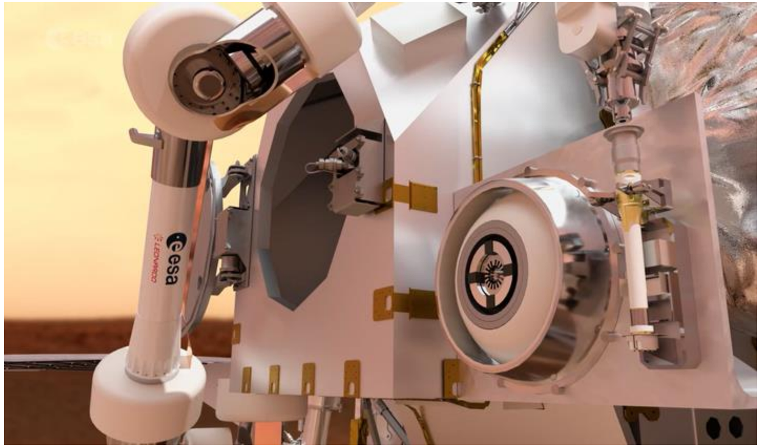
OS Lid & Workbench Localization

- The localization of both the OS Lid and Workbench will be performed using visual markers, most likely AprilTags.

- The type and disposition of markers is being defined by NASA/JPL, considering the available space and the accuracy requirements.



Credit: ESA

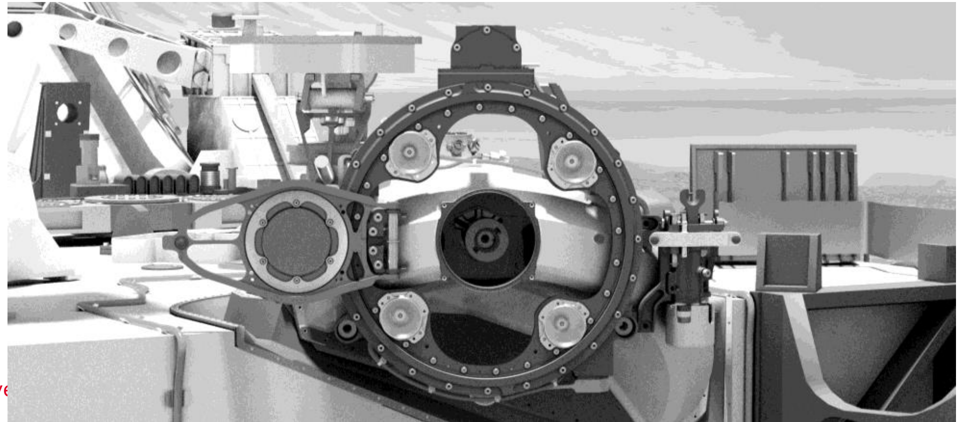


Perseverance Localization Results

- Errors for each component in Camera Frame at different distances (Far, Intermediate and Close Viewpoints)
  - The distances are given from the end-effector
  - The camera is placed backwards 27 cm from the end effector.

- The largest translation errors are obtained in the Z component (depth) and the largest rotation errors are obtained in the X and Y components
  - Typical behaviour of 3D-from-2D problems where there is much higher observability in the camera plane than in the depth

Component	50cm	10cm	3cm
Trans X [mm]	0.266	0.065	0.049
Trans Y [mm]	0.412	0.155	0.140
Trans Z [mm]	0.754	0.328	0.412
Trans Mag [mm]	**0.830**	**0.331**	**0.409**
Rot X [deg]	0.234	0.104	0.067
Rot Y [deg]	0.121	0.049	0.189
Rot Z [deg]	0.043	0.024	0.022
Rot Mag [deg]	**0.414**	**0.287**	**0.201**



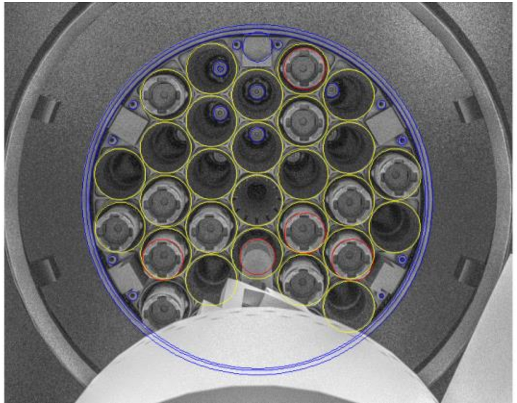
OS Localization

- Largest translation errors in the Z component, largest rotation errors in the X and Y components

- Errors are larger than in the Bit Carousel localization:
  - The camera is further from the target (~20 cm further)
  - The OS is smaller than the Bit Carousel: features are more concentrated in the centre of the image, reducing the observability

- The occupancy level of the OS (whether is empty or full of tubes) has little effect on the accuracy (results provided for the worst-case configuration)

Component	50cm	10cm	1cm
Trans X [mm]	0.234	0.042	0.035
Trans Y [mm]	0.330	0.200	0.089
Trans Z [mm]	6.557	1.261	0.532
Trans Mag [mm]	**6.501**	**1.240**	**1.263**
Rot X [deg]	2.835	0.733	0.690
Rot Y [deg]	2.246	1.547	0.775
Rot Z [deg]	0.173	0.092	0.109
Rot Mag [deg]	**4.861**	**1.659**	**1.164**



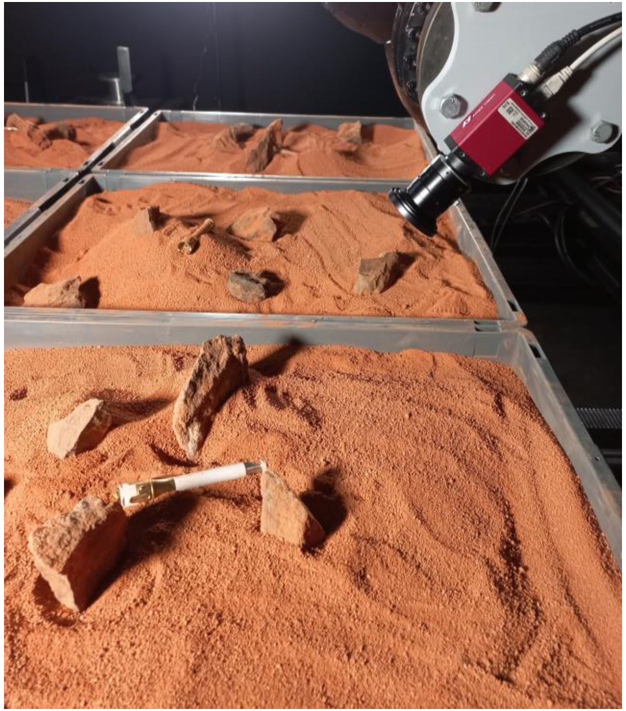
Sample Tube Localization

- The segmentation step has been widely tested using both synthetic and real imagery

- Real images have been generated using:
  - A visually representative tube model provided by NASA/JPL
  - A COTS camera with equivalent characteristics (FOV and sensor) to the flight STA camera
  - Different sandboxes replicating the appearance of the Mars terrain

- The 2D registration of the mask with the reference model resulted in accuracies better than **1mm and 0.4 degrees** when observing the tube from approximately 40 cm from the camera

- Testing of the 3D Localization is still an on-going activity



Conclusions

- The algorithms have been specifically developed to take advantage of the unique characteristics of each of the elements to be localized:
  - Perseverance Bit Carousel: the fasteners were used as relevant features to be matched against a reference model and solve the PnP problem.
  - OS container: similar approach but in this case, the circular shape of the tube slots was used.
  - Sample Tubes on the terrain: a multi-view strategy was followed to cope with the lower observability of the depth and inclination of a monocular observation
  - OS Lid and workbench: AprilTag markers proposed (under assessment)

- Achieved accuracies are in all cases compatible with the requirements of autonomy for the STA operations.

Future Work

- Validation with real mock-ups and a COTS camera with similar characteristics to the one mounted on the STA is already foreseen in the short future.
  - Both mock-ups of the Bit Carousel and OS will be provided by NASA/JPL being visually almost equivalent to the flight models, to ensure the representativeness of the tests.

- The proposed algorithms, integrated in a dedicated EGSE replicating the lander processor and operating the arm, will be used for their validation in Europe before delivery to NASA/JPL for their integration in the lander

Thank you