

Martian Fetch: Finding and Retrieving Sample-Tubes on the Surface of Mars

Jeremie Papon, Renaud Detry, Peter Vieira, Sawyer Brooks,
Thirupathi Srinivasan, Ariel Peterson, and Eric Kulczycki
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109
818-354-4729

papon@jpl.nasa.gov, detry@jpl.nasa.gov, peter.vieira@jpl.nasa.gov, sawyer.brooks@jpl.nasa.gov,
thirupathi.srinivasan@jpl.nasa.gov, arielpeterson@ucla.edu, eric.a.kulczycki@jpl.nasa.gov

Abstract—Mars Sample Return (MSR) was identified by the 2011 planetary science decadal survey as a high priority long-term goal for NASA. A three-mission campaign concept is currently being investigated. The Mars 2020 rover mission is intended to core and collect samples. These samples will be sealed in tubes and left on the surface for potential return to Earth. In the current MSR campaign concept, a Sample Retrieval and Launch (SRL) mission would collect the sample tubes left by the Mars 2020 rover and load them into a Mars Ascent Vehicle (MAV) to be launched into orbit. The third mission concept involves a spacecraft capturing the samples in Martian orbit and returning them to Earth. This paper focuses on the SRL mission concept to collect the sample tubes, addressing the problem of autonomously detecting, localizing, and grasping sample tubes deposited on the Martian surface. We employ two approaches: The first one is context-based. It would use a high precision map computed from images captured during tube release, to locate the tubes without directly observing them. The second approach directly detects the sample tubes visually and estimates their 6-DoF pose onboard from dense stereo data.

(MAV). The MAV would then release the OS into Martian orbit. The OS could then be collected by a third mission, a Sample Return Orbiter (SRO), which would capture the OS in Martian orbit before returning it to Earth. Once on Earth, these samples would enable the scientific community to use all of the resources present in sophisticated terrestrial laboratories to answer fundamental questions concerning the history and current state of Mars. In particular, returning pristine samples of Martian soil and rock to Earth will allow investigation into the past or present existence of life on Mars that would not be possible otherwise.

In this paper, we describe work towards the first step of the SRL mission concept, specifically, the task of finding and retrieving the sample-tubes deposited on the Martian surface by the first mission. The work presented is a continuation of the efforts to robustly and repeatedly localize and acquire a sample tube from a Mars-like environment for the MSR mission [1]. While that work served as an initial overall proof-of-concept, here we focus on the challenges particular to the task of finding the sample-tubes deposited on Mars. In particular, during their stay on the Martian surface, high winds and the fine-granularity of surface particles mean that sample tubes may become buried under dust. To address this, we have developed a sparse feature matching approach to localize the SRL rover with respect to images taken by the Mars 2020 rover. This gives us a pose in our 3D map, which contains a precomputed (i.e., on Earth) pose for the sample tube, allowing us to execute a "blind" grasp of the sample tube. While this permits grasping without having to actually find the sample-tube itself, it assumes that the sample-tube has not moved (e.g., from wind, or interaction with a rover).

To address this, alongside the "blind" approach, we have developed a method for localizing the tubes by direct observation. To do this, we employ a deep fully-convolutional neural network trained to detect and segment the sample tubes as well as their constituent parts (shank, bearing race, and body). This network was first trained on thousands of synthetically generated training images of dust-covered tubes, and then subsequently fine-tuned on a relatively small amount of hand-annotated real training data. In both approaches, the end result is a 6-DoF pose of the sample-tube, which we use for grasping, and ultimately, caching.

TABLE OF CONTENTS

1. INTRODUCTION.....	1
2. MISSION PROFILE.....	1
3. PLATFORM DESCRIPTION.....	2
4. VISUAL LOCALIZATION APPROACH.....	3
5. DIRECT OBSERVATION APPROACH.....	5
6. CAMERA-TO-ARM CALIBRATION AND GRASPING.....	6
7. SUMMARY AND CONCLUSION.....	7
ACKNOWLEDGMENTS.....	8
REFERENCES.....	8
BIOGRAPHY.....	9

1. INTRODUCTION

Mars Sample Return (MSR) was identified by the 2011 planetary science decadal survey as a high priority long-term goal for NASA. The Jet Propulsion Laboratory (JPL) is currently investigating a three-mission concept for an MSR campaign. As a first step in this campaign, the Mars 2020 rover, will core and collect about 31 samples. These samples will be sealed in sample-tubes and deposited on the Martian surface for potential return to Earth. Later, a second mission, currently known as the Sample Retrieval and Launch (SRL) mission concept, would collect the sample tubes left by the Mars 2020 rover (or another rover) and load them into an Orbiting Sample (OS) payload in a Mars Ascent Vehicle

2. MISSION PROFILE

The current mission profile contains two competing concepts: the first one is a mobile MAV concept which consists of a single, mobile Mars Science Laboratory (MSL) sized rover topped with the MAV, and the second one is a fetch-rover

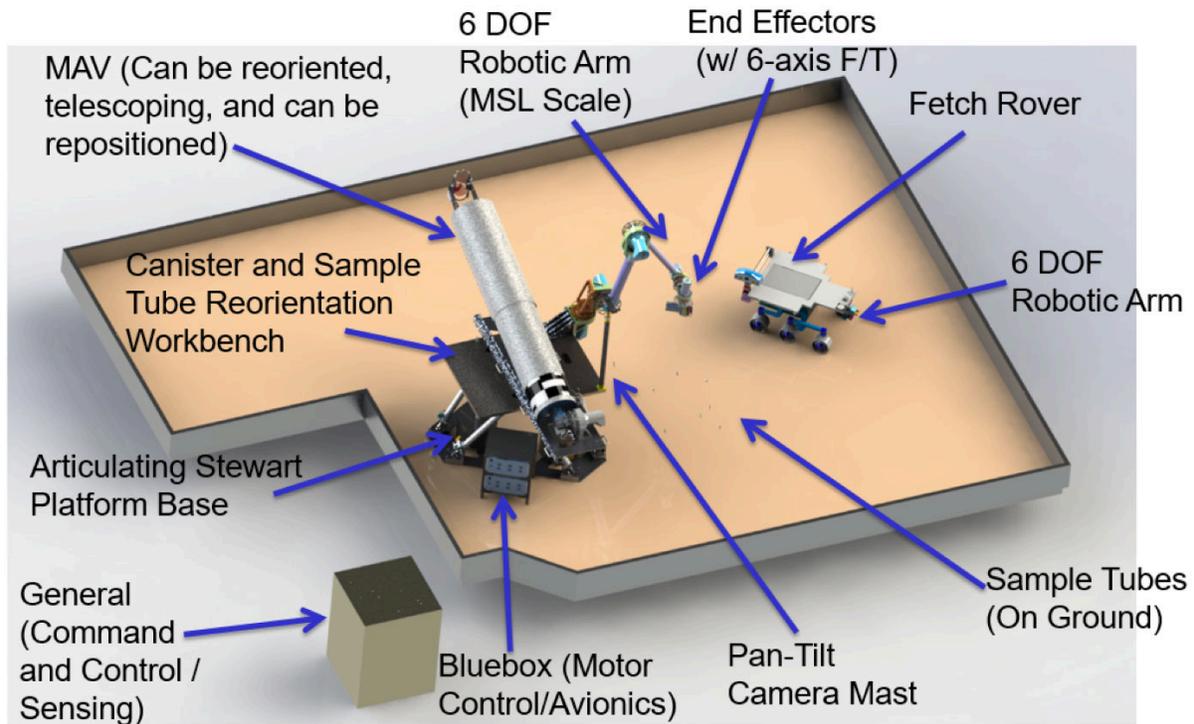


Figure 1. Mars Sample Transfer Testbed (MSTT) Infrastructure Layout.

concept which consists of a stationary MAV lander and a smaller MER-class “fetch” rover for retrieving sample tubes. This paper studies elements of both approaches via a testbed called the Mars Sample Transfer Testbed (MSTT), shown in Figure 1. The larger “SRL” testbed serves to test both the stationary MAV lander and the mobile MAV rover concepts and is based on MSL heritage. A mobility system was not developed for this testbed since mobility of this class of rover has already been demonstrated on MSL. Instead, a Gough-Stewart platform is used to simulate mobility. This testbed is an evolution of the 5 DoF brassboard arm testbed used for initial assessment of autonomous sample tube acquisition [2].

In both mission concepts, the goal for our purposes is the same – to locate sample tubes deposited on the surface, estimate their 6-DoF pose, grasp them using a robotic arm, and deposit them in a cache contained on the Mobile MAV or Fetch rover decks. While this could, in principle, be accomplished with a human-in-the-loop, doing so would add considerable length to the duration of the mission.

To illustrate this point, we can use the current MSL drilling procedure as an example. This consists of several steps: 1. Capturing Navcam images to select a drill site, 2. Selecting a drill target and perform triage observations 3. Drilling, and 4. Verifying and retrieving sample. Since each of these steps requires human verification and intervention, the operation requires a minimum of 5 Martian Sols to execute (commands to the rover are only sent once per Sol).

If one assumes that retrieving a sample tube requires at least as much intervention as a drilling campaign, the current mission profile of retrieving 31 sample-tubes will require nearly half a year. When considering the additional time required for driving between the samples, a human-in-the-loop acquisition system greatly limits the amount of time that could be spent on other science. With this in mind, the objec-

tive of this work and of the MSTT is to demonstrate that tube retrieval can be performed autonomously. If the human-in-the-loop requirement can be removed, then sample-tubes can potentially be recovered in a single Sol, greatly accelerating the SRL mission timeline.

3. PLATFORM DESCRIPTION

While both the Mobile-MAV and the Fetch-rover concepts share the same general software architecture, there are necessarily some differences in hardware and scale. For the purposes of completeness, in this Section we will briefly describe the two platforms, highlighting the differences between them relevant to the task of sample-tube retrieval.

Mobile-MAV

The Mobile-MAV architecture we have explored would be an MSL-sized rover carrying a MAV in a horizontal or near-horizontal position. It would pick up sample tubes using a gripper on the end of a 2-to-3 meter long robotic arm with between 5 and 7 degrees-of-freedom. The current SRL testbed uses a 6 DoF arm with a 1-DoF parallel jaw gripper for sample tube acquisition. The same arm would reorient the tubes (if necessary), load the individual tubes into a sample canister, and ultimately load the sample canister into the MAV. The rover would have several cameras which could be used to identify and locate the tubes, including a Navcam (a stereo-pair mounted on a 2-meter tall pan/tilt mast), Hazcams (a fixed stereo-pair mounted on the front of the rover at approximately 1.0m above the surface), and a Toolcam (a single camera at the end of the arm).

Fetch Rover

The Fetch-rover we used was developed at JPL 18 years ago as a prototype for the Mars Exploration Rovers mission. This



Figure 2. Images captured by the left mast camera. From left to right, top to bottom: board configurations 1, 5, 11 and 16.

battery-powered, untethered, Mars Exploration Rover (MER) sized rover was successfully used for coring and sample caching experiments in Mono Lake in 2010, in support of the future Mars 2020 mission. The rover avionics and software were updated to support the MSTT task. The mobility system uses a rocker-bogie suspension with six wheels, each with drive and steering actuators. The avionics of the Fetch-rover include a tethered or battery powered power system, control and perception computers, hardware drivers and a camera suite. Finally there is a 6-DoF arm at the front of the rover with a 6-axis force sensor and a 1-DoF three fingered gripper on the end. The Fetch-rover has a Navcam and a Hazcam like the Mobile-MAV, but does not currently have a Toolcam, due to space and weight constraints of the arm and end-effector.

4. VISUAL LOCALIZATION APPROACH

In its current form, the MSR campaign anticipates a delay of several years between releasing the tubes with the Mars 2020 rover and recovering them with the SRL rover. The Martian atmosphere is meager by comparison to Earth's – its average atmospheric pressure at ground level is only 0.6% of Earth's mean sea level pressure. Mars' geology is also much less active than Earth's. Mars landers and rovers have accumulated evidence showing that under those conditions rocks larger than a few centimeters are unaffected by weather: winds are not strong enough to move those rocks, and although a small dune is likely occasionally build up at their base, the rocks are unlikely to become covered in sand. From those observations, we hypothesize that (1) sample tubes will not move between release and recovery, but that (2) there is a substantial risk of a dune building up on the sides of the tubes, and of the tubes becoming covered with a thin layer of dust. Under the second hypothesis, detecting the tubes via direct observation may not be possible.

This section addresses the case of recovering tubes that are partly covered in sand and that cannot be detected by direct observation. Instead, we propose to use images taken by the Mars 2020 rover to build a sparse feature map encoding a tube's location relative to nearby landmarks and ground

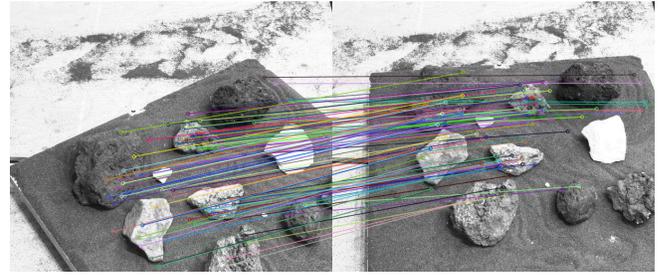


Figure 3. Features matched between configuration 4 (left) and configuration 14 (right).

topology, and let the SRL rover localize the tube by registering the map to the terrain.

We assume that the Mars 2020 rover selects release locations that contain several visual landmarks, such as larger rocks or hard terrain features. We also assume that when the SRL rover can be driven to within a meter of the tube's location via global positioning. We address the problem as follows: Upon releasing a tube, the Mars 2020 rover captures a stereo image showing the tube and its direct surroundings. An operator subsequently annotates those images, labeling areas that are unlikely to change over the next ten years, large rocks for instance. The operator then runs a feature detector in the labeled areas, and pairs features that are similar in appearance and that respect epipolar constraints. The operator also computes by inspection the pose of the tube in the camera frame. Upon reaching the same site some years later, the SRL rover extracts a set of features from a stereo view of the site, and searches for four-way feature correspondences, i.e., quadruples of features SRL-left, SRL-right, Mars-2020-left, Mars-2020-right. Quadruples that do not respect SRL's epipolar constraints are pruned, and the rover computes the Mars-2020-to-SRL transformation that minimizes the reprojection error of all features.

The remainder of this section presents a proof-of-concept experiment that quantifies the accuracy of visual pose estimation in a Mars-like environment, with cameras that are representative of upcoming Mars missions. The conclusion of this experiment (detailed below) is that the Mars 2020 camera hardware allowed us to register pairs of stereo images captured from viewpoints that are up to 1m apart, with up to 25° orientation differential, with a 1.5% mean relative error in translation and a 3.1% mean relative error in rotation.

The testbed stereo camera (Figure 1) is attached to a fixed base. To simulate different robot poses, we arranged ten rocks on a 1.15×1.15m board covered with sand, and moved the board to 16 different configurations within the field of view of the camera, capturing a stereo image pair in each configuration (Figure 2). We extracted ground-truth poses with a VICON motion-capture system, via four IR markers attached to the sides of the board. We then applied a pose estimation algorithm (described below) to the 256 different pairings of the 16 board poses.

Our pose estimation algorithm follows the approach of Geiger et al. [3] and relies on the LIBVISO2 codebase.² The algorithm takes as input two pairs of rectified stereo images, and the camera's intrinsic and extrinsic parameters. It works by extracting features from the images, matching the features,

²<http://www.cvlibs.net/software/libviso/>

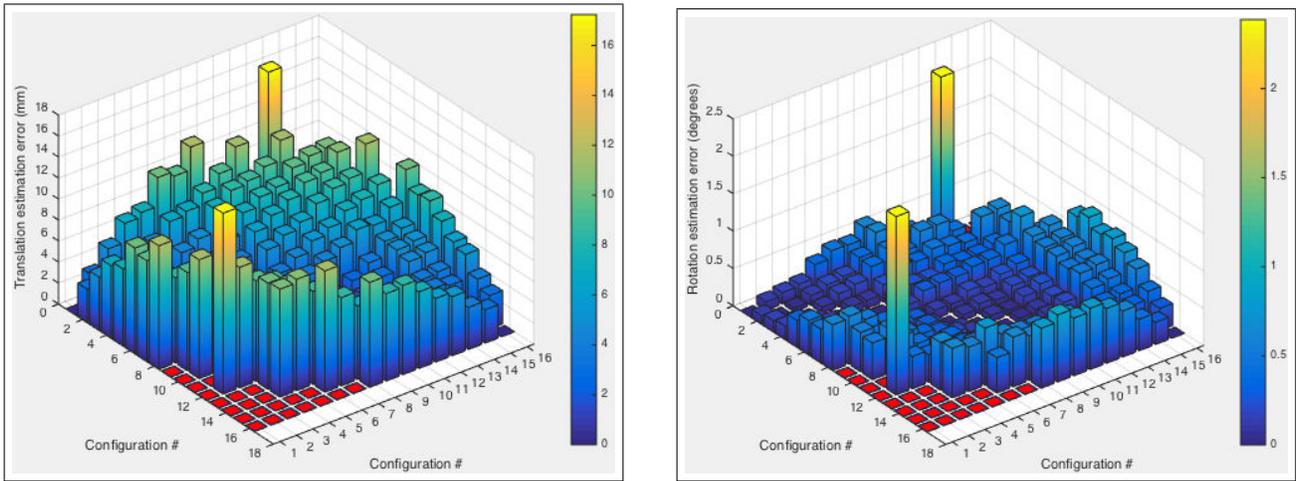


Figure 4. Translation and rotation errors for different offsets in camera viewpoints. Each cell (i, j) shows the error between ground truth pose j , and ground truth pose i transformed with the vision-based $i \rightarrow j$ estimated transformation. Red cells indicate that the number of matches was insufficient to produce a reliable estimate.

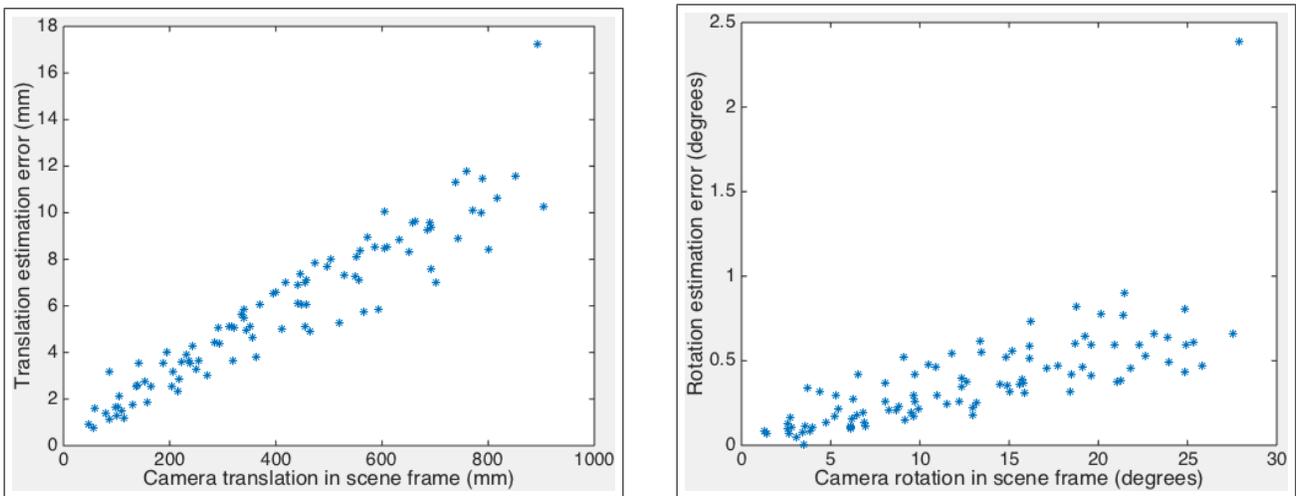


Figure 5. Translation and rotation errors as a function of translation and rotation offsets between Mars 2020 and SRL camera viewpoints. As one would expect, increasing the rotation and translation offset between the two viewpoints results in increasing amounts of error.

and finally computing the transformation that minimizes the feature reprojection error. We compute features by filtering the input images with a 5×5 corner mask and with a blob mask of the same size, followed by non-maximum- and non-minimum-suppression. We then compute four-way feature correspondences, and prune those that do not respect the epipolar constraints of either stereo pair. As the moving board only covers a fraction of the field of view of the cameras, we filter out features coming from the scene background, using a mask drawn by hand for each image. We note that this artifact puts our experiment at a slight disadvantage compared to the mission scenario, as only a fraction of the field of view is exploited for motion estimation. To compute the transformation between the two board configurations, we minimize the reprojection error of all features, first via RANSAC, then with Gauss-Newton on the RANSAC inliers. Figure 3 shows matching features across the left images of board configurations 4 and 14.

To quantify the reliability of pose estimation, we compared

the vision-based relative transformations to those acquired with the motion-capture system. Figures 4 and 5 show the translation and rotation errors between estimates and ground truth. Figure 4 shows the translation (left) and rotation (right) errors for all pairs of board configurations. Red squares correspond to cases for which the algorithm could not find a convincing solution. The rightmost graph of Figure 4 shows larger errors for cases involving configurations 14, 15 and 16. Those errors are explained by the fact that the corresponding images only show a part of the board, as it is sliding out of the field of view. Case 2–12 is a border case, with only 15 challenging four-way correspondences. Figure 5 shows the same errors, as a function of the camera translation/rotation that would have occurred if the board had been fixed – instead of showing errors as functions of the board’s translation/rotation in the camera frame. We opted for these plots because they are representative of the scenario we are characterizing and can be interpreted as the pose differential between Mars 2020 and SRL. The mean position error is 1.5% (relative to hypothetical camera translation) and

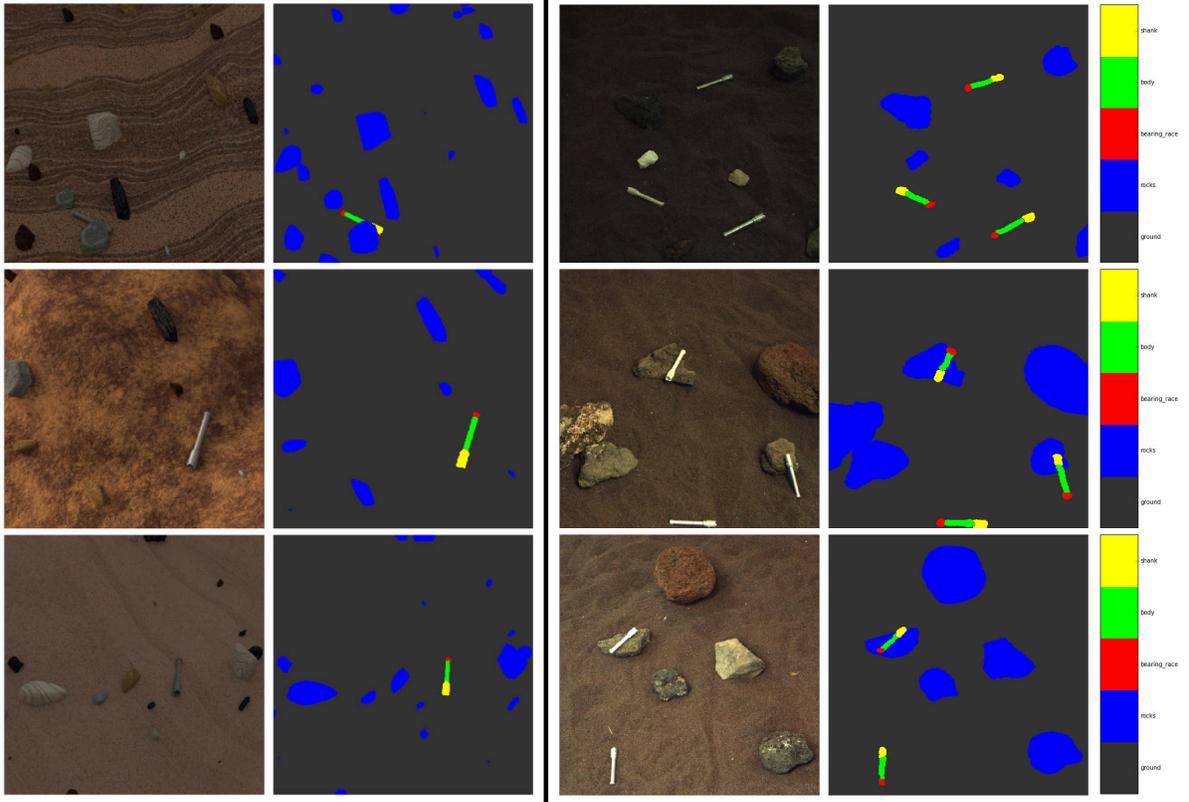


Figure 6. Examples of synthetic (left) and real (right) training data and labels generated to train the deep FCNN. Synthetic data was generated using the Blender and physics simulator. Real data was captured in our lab and in the Mars Yard at JPL.

the mean orientation error is 3.1% (relative to hypothetical camera rotation).

5. DIRECT OBSERVATION APPROACH

As an alternative to the "blind" grasping approach described above, we also developed a direct observation procedure which aims to localize sample tubes using only the current stereo-camera observations. This procedure is considerably more difficult than the blind approach, as it can only use the features on the tube itself for detection and pose estimation - features which are likely to be occluded by dust and other debris. Furthermore, the tube itself has a high degree of symmetry, meaning that even when properly detected, there are often multiple possible poses which fit the observations equally well.

Our direct observation approach consists of two primary steps: detection and pose estimation. The first step aims to detect the sample tubes in either Navcam or Hazcam imagery and then extract the pixels belonging to them. The second uses the extracted pixels in order to estimate the 6-DoF pose of the sample tube in the rover's reference frame so that a proper grasp approach trajectory can be determined.

For sample tube detection and segmentation, we employ a type of deep Convolutional Neural Network (CNN) known as a Fully-Convolutional Network (FCNN) [4]. FCNN's do away with the fully-connected layers typically used in the later stages, and instead only use convolutional layers, allowing end-to-end training from RGB-image directly to semantic

label space. In our case, we use a novel residual-dilate-skip architecture 7 to provide an accurate per-pixel label in a single forward pass. This network architecture takes advantage of recent work in residual connections [5] and dilated layers [6] to increase understanding of the relationships between parts. Additionally, we use a skip-network architecture to provide sharp object boundaries and avoid the over-smoothing effects commonly seen in simpler upsampling schemes. Our network is trained to localize the sample tubes as well as their constituent parts (shank, bearing race, and body). Localization of parts is especially helpful in disambiguating symmetries in pose.

As this is a highly specialized application, an annotated dataset was not readily available for training. As such, we decided to use simulation to create an initial training set, as proposed in [7]. The choice to use simulated data was based on three factors: First, there exist no examples of what a sample tube that has been Mars for several years will look like. Second, the design of the sample tube has not been finalized, and it is relatively inexpensive to regenerate our synthetic training data when the design changes. Finally, the amount of data required to train a deep network is significant, and we simply did not have the resources to collect and hand-annotate the amount of data necessary to train a network from scratch.

For generation of simulated data we used the open-source Blender [8] API, allowing us to generate random scenes programmatically. Our data set consists of 50,000 randomly generated scenes in which sample tubes were dropped from a height of two meters, allowed to settle, and then subsequently

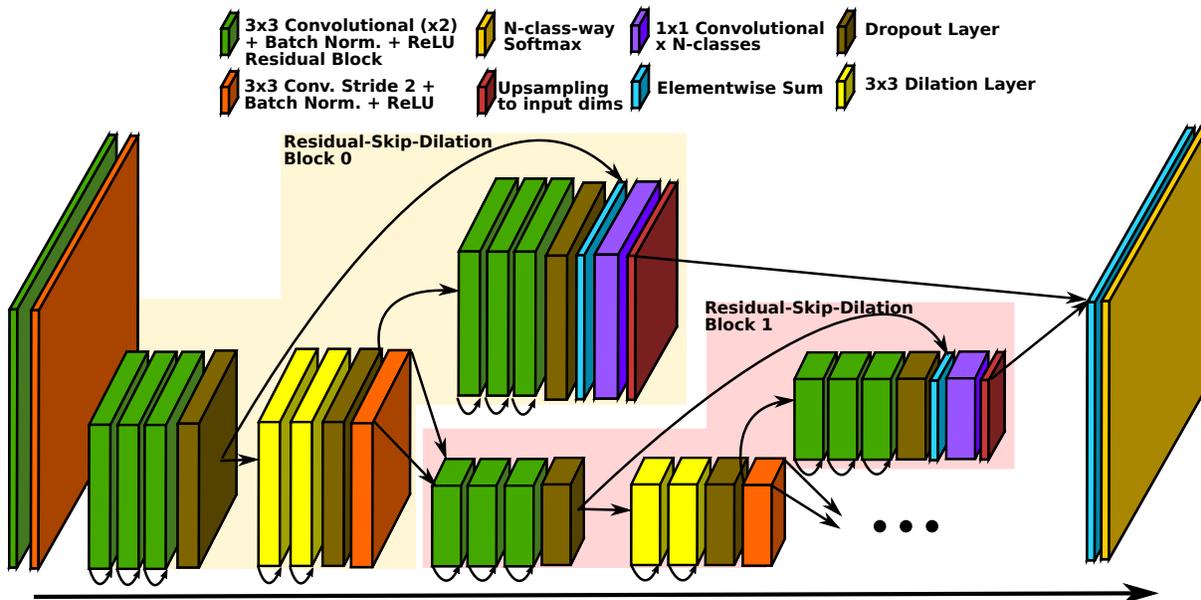


Figure 7. Residual-Skip-Dilation architecture used to detect and segment out sample tubes. Any number of Residual-Skip-Dilation blocks can be used. In this work, we used two, finding that to be sufficient, since spatial context was only useful near the sample-tube for distinguishing between the sample-tube parts.

coated with a simulated layer of dust. Additionally, location and intensity of illumination and surface conditions were varied for each render. Examples of our synthetic training data can be seen in Figure 6.

Once the network was trained on the synthetic data, it was subsequently fine-tuned on a relatively small (150 images) amount of hand-annotated real training data collected in our indoor lab and in the outdoors Mars Yard at JPL. Examples of these images, as well as the hand-annotated labels can be seen in Figure 6. While the real data looks somewhat different than our synthetic scenes, especially in the sand and rocks, the synthetic data is probably sufficient to begin learning the overall shape and relative positioning of parts of the sample-tubes.

The output of our fully convolutional network is a per-pixel labeling of ground, rocks, and sample-tube, split into three parts - bearing-race, shank, and body. Examples of this classification output are shown in Figure 8. The sample tube was split into three parts in order to prevent pose estimation errors arising from 180-degree flips in yaw. For pose-estimation, the pixels classified as sample-tube are extracted and mapped into the 3D pointcloud reconstructed from either the Navcam or Hazcam stereo disparity. The 3D points corresponding to sample tubes are then partitioned using Euclidean Clustering [9].

Once we have a number of candidate clusters of 3D points, we then attempt to fit the sample tube model to each cluster using Generalized Iterative Closest Point [10], providing a 6-DoF pose in the camera reference frame. The cluster with the best fit (in terms of point-to-point distance) is then selected for grasping.

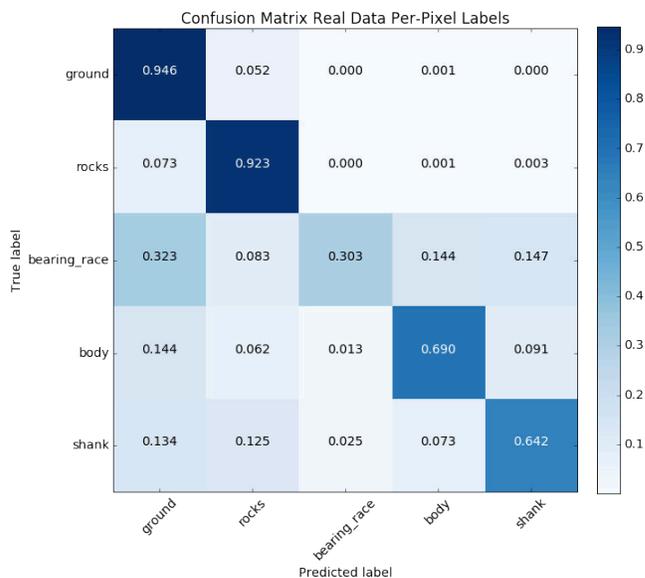


Figure 9. Confusion Matrix of per-pixel labeling accuracy for the five classes.

6. CAMERA-TO-ARM CALIBRATION AND GRASPING

Once the pose of the tube has been determined using either the blind or the direct approach, we must transform the pose from the camera reference frame to the arm gripper tool frame. As any error in the initial grasp will propagate through the rest of the manipulation chain, possibly causing errors later on, we must grasp the tube as near to our goal position as possible. With this in mind, we adopt a two-step camera-to-arm calibration procedure: First, we move the arm to a standard starting position above the workspace, and capture an image of a fiducial marker (an April tag [11]) on the end-

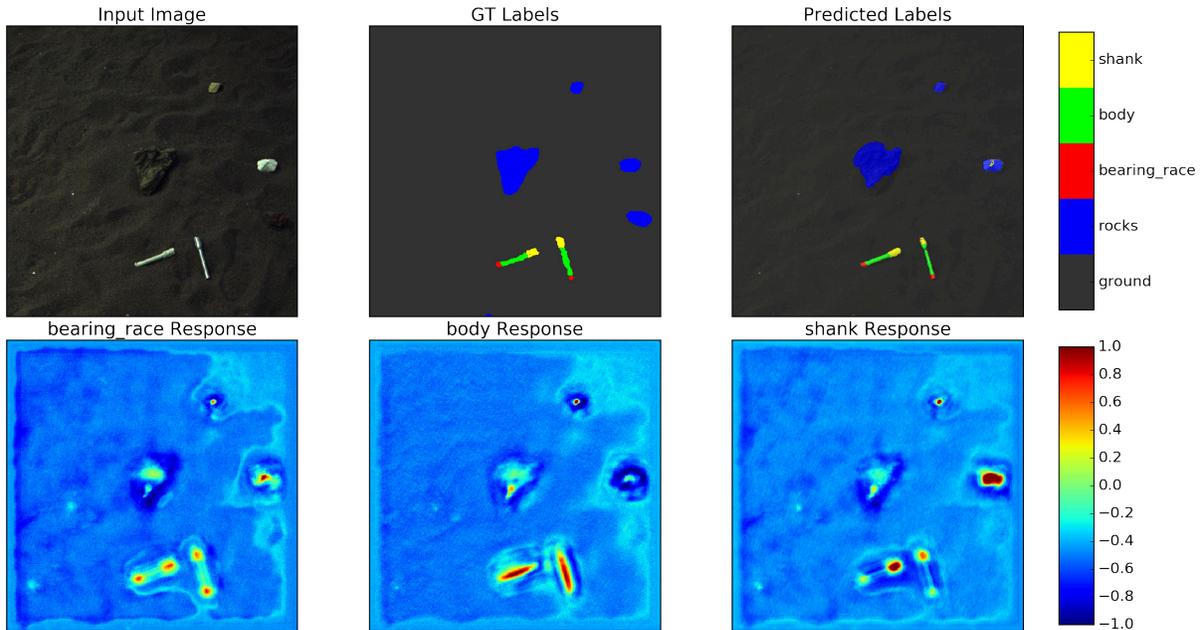


Figure 8. Example of classifier labeling output on real data showing the activations from the final convolutional layer after soft-max.

effector - see Figure10 - allowing us to calculate a camera-to-arm transformation. Next, the manipulation software moves the end-effector to a standoff pose a small distance above the detected sample tube. This standoff position must be sufficiently close to the tube to take advantage of relative accuracy instead of global accuracy, but far enough to prevent the end-effector from hitting the tube or the ground if the tube position or the camera-to-arm calibration is inaccurate. Once the end-effector has reached the standoff position, we capture another image of the fiducial marker, and re-estimate the camera-to-arm transformation. This allows us to correct for arm deflection as well as pointing error in the Navcam pan/tilt unit. We found that this second step was essential, sometimes resulting in a correction of over 1cm in grasp position.

Finally, using our corrected camera-to-arm transformation, we select a grasp approach orthogonal to both the tube's principal axis and the horizontal axis. This ensures that the gripper can grab the tube in an orthogonal ('body') grip, and generally prevents end-effector collisions with the terrain. Assuming the tube location is within the workspace of the robotic arm, we move the gripper in a straight-line trajectory from its original standoff position to a much smaller standoff position (1cm) above the refined tube position. Finally, we use force sensing to contact the tube, and then drive the gripper to stall around the tube. These last two steps are able to accommodate a small amount of position error along the axes orthogonal to the tube. While this is only the beginning of the sequence needed to retrieve and cache the sample-tubes, the rest of the chain does not yet use computer vision, and is beyond the scope of this paper.

7. SUMMARY AND CONCLUSION

Autonomously recovering sample tubes dropped on the Martian surface is a difficult task, but one that is necessary for recovering a large number of samples within a reasonable amount of time. In this work, we have presented two solutions to the problem; one that uses localization and pre-

computed sample-tube poses, and another that attempts to detect and estimate the pose of the sample-tubes directly on the rover. While we have shown that both of these methods are effective for sample-tube recovery, each is suited to a slightly different application.

The blind recovery method is well suited to recovering sample-tubes which are either partially or fully buried, as it does not require any actual observation of the sample-tube itself. Unfortunately, this relies on the fact that the sample-tube has not been displaced from the position it was in immediately after it was dropped. Should the recovery rover happen to move the sample-tube accidentally (e.g. due to a failed grasp attempt, or during driving), the pre-computed tube pose will no longer be valid, and the method will fail.

Conversely, the direct-observation method has no hard assumptions about the position of the tube. This means that recovery from failed grasping is possible; in fact, one could even envision a system which uses an air-gun or a brush to clean the sample-tube of dust before grasping it. Of course, this method relies on the neural-network's ability to detect the tube autonomously. While this is achievable if the tube is readily visible, it might not be possible under significant dust accumulation.

In the end, it is likely that in practice a combination of both methodologies will be necessary for robust localization and grasping of sample-tubes. Indeed, the safest route is likely to use both methods, and only attempt an autonomous grasp when they agree closely, falling back to human-in-the-loop control when they do not. In fact, this may allow for the strengths of each method to be leveraged to help the other - the blind method can be used to greatly constrain the search space for the direct approach, while the direct approach can be used to correct small displacements of the tube.

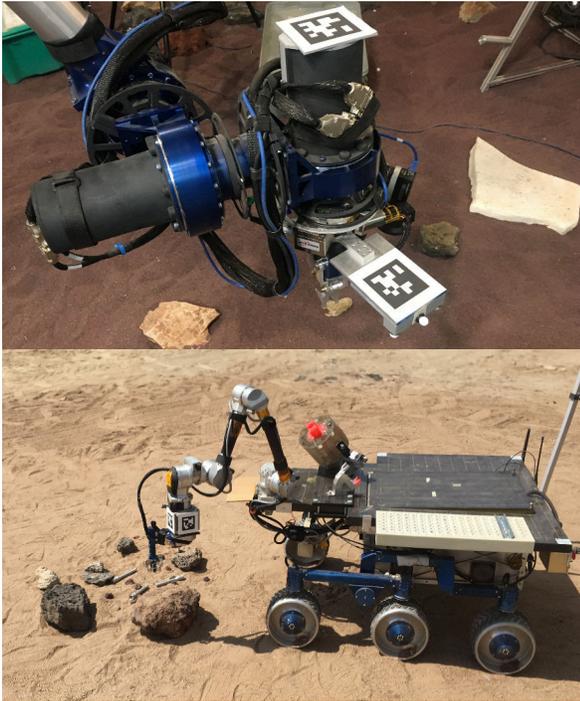


Figure 10. Both platforms have April-tag fiducial markers mounted on the last joint. This permits online refinement of the camera-to-arm calibration, greatly improving grasp accuracy.

ACKNOWLEDGMENTS

The authors wish to thank all of the members of the Mars Sample Transfer Testbed (MSTT) team for their long hours, hard work, and dedication to testing and improving the testbed and rover capabilities. A special thanks to Chris Turner for all his time spent creating the real sample-tube dataset, which made all of this possible, and to Dennis Wai, Christina Kneis, and Jasmine Moreno for extensive operation of the STARM and Pluto testbeds, which validated these approaches with real hardware. Finally, we would like to acknowledge Joe Parrish of the JPL Mars Program Formulation Office for funding and directing this work. The research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

REFERENCES

- [1] R. Mattingly and L. May, "Mars sample return as a campaign," in *Aerospace Conference, 2011 IEEE*. IEEE, 2011, pp. 1–13.
- [2] K. Edelberg, J. Reid, R. McCormick, L. DuCharme, P. Backes, and E. Kulczycki, "Autonomous localization and acquisition of a sample tube for mars sample return," in *AIAA SPACE 2015 Conference and Exposition, 2015*, p. 4483.
- [3] A. Geiger, J. Ziegler, and C. Stiller, "Stereoscan: Dense 3d reconstruction in real-time," in *Intelligent Vehicles Symposium (IV)*, 2011.
- [4] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and*

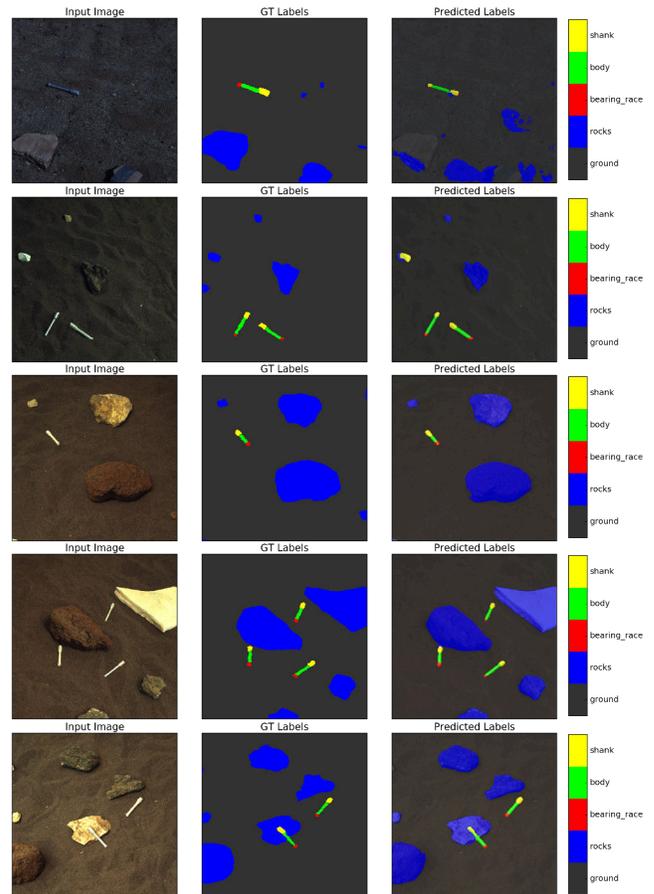


Figure 11. Examples of labeling on data collected in Mars Yard and from our test-bed. Hand annotated labels are shown in the center, with network labeling outputs in the right column.

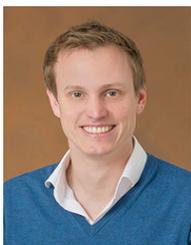
Pattern Recognition, 2015, pp. 3431–3440.

- [5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *arXiv preprint arXiv:1512.03385*, 2015.
- [6] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," *arXiv preprint arXiv:1511.07122*, 2015.
- [7] J. Papon and M. Schoeler, "Semantic pose using deep networks trained on synthetic rgb-d," in *Proceedings of the IEEE International Conference on Computer Vision, 2015*, pp. 774–782.
- [8] Blender Online Community, *Blender - a 3D modelling and rendering package*, Blender Foundation, Blender Institute, Amsterdam, 2016. [Online]. Available: <http://www.blender.org>
- [9] R. B. Rusu and S. Cousins, "3d is here: Point cloud library (pcl)," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*. IEEE, 2011, pp. 1–4.
- [10] A. Segal, D. Haehnel, and S. Thrun, "Generalized-icp," in *Robotics: Science and Systems*, vol. 2, no. 4, 2009.
- [11] E. Olson, "AprilTag: A robust and flexible visual fiducial system," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, May 2011, pp. 3400–3407.

BIOGRAPHY



Jeremie Papon received his B.Sc in Electrical Engineering from the U.S. Naval Academy in 2007 and his M.Sc degree in Electrical Engineering from Stanford University in 2009. After this he shifted continents and focus, moving to the Computational Neuroscience group at Georg-August University, Goettingen, where he received his PhD in Computer Science in 2014. He joined JPL in 2015 as a Robotics Technologist, specializing in computer vision. His research focuses on semantic understanding of images and point clouds, convolutional neural networks, sensor fusion, and localization & mapping in noisy and extreme environments.



Renaud Detry is a research scientist at NASA JPL, and a visiting researcher in the Systems and Modeling Group (University of Liège, Belgium) and in the Computer Vision and Active Perception lab (KTH Kungliga Tekniska Högskolan, Stockholm, Sweden). He earned an engineering degree at the University of Liège in 2006, and a Ph.D. in robot learning from the same university in 2010. He subsequently earned Junior Researcher starting grants from the Belgian FNRS and from the Swedish VR. He alternated between KTH Stockholm and the University of Liège between 2010 and 2015, before joining the Robotics and Mobility Section at JPL in 2016. His research interests are in perception for manipulation, robot grasping, computer vision and machine learning.



Peter Vieira joined the Robotic Manipulation & Sampling group at JPL, in August, 2014. Before JPL Pete completed his Master of Science in Electrical & Computer Engineering at Georgia Institute of Technology where he held a Graduate Research Assistantship, during which he was a main team member of their Darpa Robotics Challenge team, *Hubo*, and co-operated the robot during the semi-finals in Miami, FL. The focus of his Masters was robotics software and control systems. He gained a lot of hands-on experience with whole robotics systems.



Sawyer Brooks is a software engineer in JPL's Robotic Manipulation and Sampling Group. He received his B.S.E in Mechanical Engineering and his M.S.E in Robotics from the University of Pennsylvania in 2014 and 2015, respectively. His interest in robotics was originally sparked in 6th grade by the *Spirit* and *Opportunity* rovers, and he continues to be astonished by the fact that he now works with the designers of those robots. In addition to MSTT, he works on the first portion of Mars Sample Return as a software engineer on the Mars 2020 Sample Caching Subsystem. He also led the development of *Limbi*, a self-mobile robotic limb for in-space assembly. His research interests include bringing greater autonomy to robotics in

aerospace, robots for planetary habitat assembly, and multi-robot cooperation. In his free time, he hikes, skis, and hangglides.



Thirupathi Srinivasan received a B.S. degree in Aerospace Engineering from the California State Polytechnic University, Pomona in 2015 and is currently pursuing an M.S. in Mechanical Engineering from the University of California, Los Angeles. He is a systems engineer in JPL's Verification and Validation Group, where he focuses on developing and conducting tests for MSTT, as well as developing enterprise software applications for expanding the model-based systems engineering (MBSE) tool suite available at JPL.



Ariel Peterson is currently pursuing a B.S in Computer Science and Engineering from the University of California, Los Angeles. She has interned at JPL for the past summer working on Research and Development for a potential Mars Sample Return Campaign. She has worked on machine vision software and testing with the Mars Sample Transfer Test-bed.



Eric Kulczykcki received a dual B.S degree in Mechanical Engineering and Aeronautical Science and Engineering from the University of California, Davis, in 2004. He received his M.S. degree in Mechanical and Aeronautical Engineering also from the University of California, Davis in 2006. He is a member of the engineering staff at the Jet Propulsion Laboratory, California Institute of Technology, where he is currently involved in Mars sample transfer chain technology development, sampling technology development for extreme environments, Mars 2929 Sample Caching Subsystem, and mechanical design of various mobility platforms. He has worked at JPL for over 12 years.