Machine Vision based Sample-Tube Localization for Mars Sample Return

**Shreyansh Daftry, Barry Ridge, William Seto, Tu-Hoa Pham, Peter Ilhardt*, Gerard Maggiolino*,
Mark Van der Merwe*, Alex Brinkman, John Mayo, Eric Kulczyski and Renaud Detry**
Jet Propulsion Laboratory, California Institute of Technology
Pasadena, CA, USA
`Shreyansh.Daftry@jpl.nasa.gov`

Abstract—A potential Mars Sample Return (MSR) architecture is being jointly studied by NASA and ESA. As currently envisioned, the MSR campaign consists of a series of 3 missions: sample cache, fetch and return to Earth. In this paper, we focus on the fetch part of the MSR, and more specifically the problem of autonomously detecting and localizing sample tubes deposited on the Martian surface. Towards this end, we study two machine-vision based approaches: First, a geometry-driven approach based on template matching that uses hard-coded filters and a 3D shape model of the tube; and second, a data-driven approach based on convolutional neural networks (CNNs) and learned features. Furthermore, we present a large benchmark dataset of sample-tube images, collected in representative outdoor environments and annotated with ground truth segmentation masks and locations. The dataset was acquired systematically across different terrain, illumination conditions and dust-coverage; and benchmarking was performed to study the feasibility of each approach, their relative strengths and weaknesses, and robustness in the presence of adverse environmental conditions.
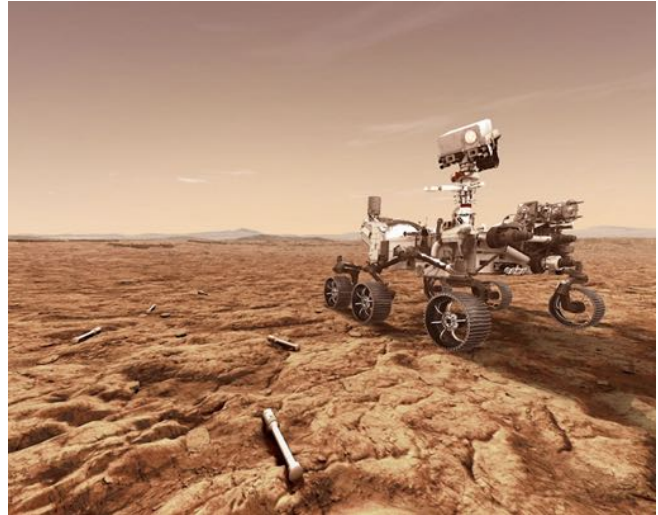
TABLE OF CONTENTS

1. INTRODUCTION

Determining the habitability of past and present martian environments continues to be the focus of current and future missions to Mars. Indeed, studying the geological history of Mars holds the key to both understanding the origins of life on Earth and in the Solar System. While recent and ongoing robotic missions have revolutionized our understanding of the red planet [1], [2], the results from orbital and in-situ surface robotic missions alone are not sufficient to fully answer the major questions about the potential for life, past climate, and

Figure 1. NASA's Mars Perseverance rover, which serves as the first stage of the MSR campaign, will store rock and soil samples in sealed tubes on the planet's surface for future missions to retrieve, as seen in this illustration.

the geological history of Mars. Even if an orbital or in situ mission were to discover putative evidence for the existence of past or present life on Mars, confirming these results would necessitate that samples be collected, returned to Earth, and verified by multiple rigorous laboratory analyses on Earth. As a result, returning samples from Mars back to Earth was identified as the highest priority planetary science objective in the Planetary Science Decadal Survey [3].

In response to that, a potential concept for a Mars Sample Return (MSR) architecture is being jointly studied by NASA and ESA [4]. As currently envisioned, the MSR campaign consists of a series of 3 missions: sample cache, fetch and return to Earth. First, NASA's Perseverance Rover, launched in 2020, will collect scientifically selected samples and store them in sealed tubes on the planet's surface (see Figure1), for possible return to Earth. Then, a potential future mission, with a Sample Retrieval Lander (SRL), would collect the sample tubes and load them into an Orbiting Sample (OS) payload in a Mars Ascent Vehicle (MAV). The MAV would release the OS into Martian orbit. The third mission, an Earth Return Orbiter (ERO), would rendezvous with the samples in Mars orbit and ferry them back to Earth.

This work focuses on the SRL mission concept [5] to collect and retrieve the sample tubes. The SRL mission would deploy a lander in the vicinity of Jezero Crater, where the Mars

Perseverance rover plans to land and collect and cache samples during its 1.25-Mars-year primary surface mission. Key payloads on SRL would include an ESA-provided Sample Fetch Rover (SFR) and Sample Transfer Arm (STA), and a NASA-provided OS and MAV. Once on the surface, the SFR would egress from SRL and begin its surface mission to retrieve samples previously cached by Mars Perseverance Rover at one or more depot locations. The solar-powered SRL mission would then carry out its surface activities during Martian spring and summer, maximizing available power, and would complete its surface mission and launch the retrieved samples into orbit before start of northern hemisphere fall, prior to significant decrease in available solar power and in advance of the potential for global dust storms. According to the currently envisioned surface mission timeline, this would allocate ~ 150 sols for SFR to complete this retrieval and return to SRL. With a nominal 4-km SRL landing dispersion, this drives the need for high levels of SFR autonomy and mobility, with support for drive distances of up to ~ 500 m/sol, and with efficient operations during tube pickup at the depot site(s) and tube transfer to SRL.

In this paper, we study the problem of autonomously detecting and localizing sample tubes deposited on the Martian surface directly from camera images. In Section 4 we discuss the first of two machine-vision approaches that we studied, a geometry-driven approach based on template matching that uses hard-coded filters and a 3D shape model of the tube; and in Section 5 we discuss the second, a data-driven approach based on convolutional neural networks (CNNs) and learned features. In Section 6, we present a large benchmark dataset of sample-tube images, collected in representative outdoor environments and annotated with ground truth segmentation masks and locations. The dataset was acquired systematically across different terrain, illumination conditions and dust-coverage. Finally, in Section 7, we describe the benchmarking evaluations that were performed using well-known metrics to compare and assess the two methods using this data.

2. RELATED WORK

Mars Sample Return

Several studies have been conducted over the last few decades with the goal of defining science objectives for Mars Sample Return [3], [6], [7], [8] and potential mission concept architectures [9], [10], [11], [12]. Furthermore, to enable this vision, several research and technology development efforts [13] have been under way - from autonomous rover technology [14], [15] to OS design and capture-systems [16], [17]. Our work is closely related to, and a continuation of, the efforts to demonstrate robust localization and autonomous retrieval of sample-tubes from a Mars-like environment for potential Mars Sample Return. While [18] served as an initial proof-of-concept, subsequent work [19], [20], [21] further studied the problem of direct and indirect sample-tube localization in an analog indoor testbed. In this work, we do a comprehensive study on the performance of sample-tube detection algorithms on a benchmark dataset collected in a representative outdoor environment.

Image-based Object Detection and Localization

Object detection and localization, as of one the most fundamental and challenging problems in machine vision, has received great attention in recent years - see [22] for a review. Traditionally, the problem of object detection and localization is tackled by matching feature points between 3D models and images [23], [24]. However, these methods require that there are rich textures on the objects in order to detect features for matching. As a result they are unable to handle texture-less objects like sample-tubes. More recently, with the advent of deep learning, CNNs by comparison, have made significant progress in object classification [25], detection [26], [27], semantic segmentation [28] and instance segmentation [29], including their application to Mars rover autonomy [30], [31].

For texture-less object detection, the taxonomy of approaches can be broadly classified into three categories: feature-based, view-based and shape-based. The feature-based approaches [32], [33], [34] match 2D image features like edges or line segments to the corresponding 3D model points. The view-based approaches [35], [36], [37] compare a given image of the object with its pre-computed 2D views. Lastly, shape-based approaches [38], [39], [40], [41], [42] are based on template matching of edge-segments. Furthermore, end-to-end learning based methods [43], [44], [45] have also been proposed that utilize deep learning for the task of 6D pose estimation directly from RGB images as the only input modality.

3. SAMPLE-TUBE LOCALIZATION

The operational context assumed is as follows. The Perseverance rover will place sample tubes on the ground at one or multiple (presently-unknown) sites referred to as *tube depots*. Within each tube depot, the sample-tubes will be placed several meters apart in areas where sand is not abundant and that have a relatively flat ground with slopes less than $25°$. It is to be noted that Perseverance has little control over precise tube placement: tubes will be released from the rover's underbelly, and may bounce and roll on the ground before coming to a halt. Furthermore, it is anticipated that SFR is capable of autonomously driving from tube to tube – a capability addressed in a separate paper [46]. For the context of this paper, this implies that SFR can drive itself to a pose that places a tube within its field of view.

Based on these assumptions, our understanding of the Martian conditions (weather, geology, etc.) and through qualitative observation of past lander and rover images, we can conservatively hypothesize the following constraints: (a) tubes will not move, (b) sand or dust may, with low probability, pile up next to tube or rocks, forming drifts, and (c) dust will deposit everywhere, creating a dust layer that will not exceed $0.25\,\mathrm{mm}$ in thickness. Rovers typically operate between 10am and 4pm, yielding a large range of lighting conditions, including shadows incurred by rocks or by the rover itself.

Our goal is to then to study the robustness for image-based detection and localization of sample-tubes under these constraints. More specifically, we quantify the performance of two object localizers for sample-tube detection in a variety of of conditions that include clear, non-occluded tubes, partial occlusions by rocks or sand, or partial or complete shadows. We focus on the detection and localization only from a single images for two reasons: first, it allows the proposed algorithms to be applicable to a diverse set of scenarios (e.g. if we want to detect using either a mast-mounted stereo camera or a wrist-mounted monocular camera), and second, based on our preliminary analysis we expect depth data on the sample-tubes to be noisy and thereby not very helpful for the detection task. Extensive analysis on the usefulness of stereo range data is beyond the scope of this paper.
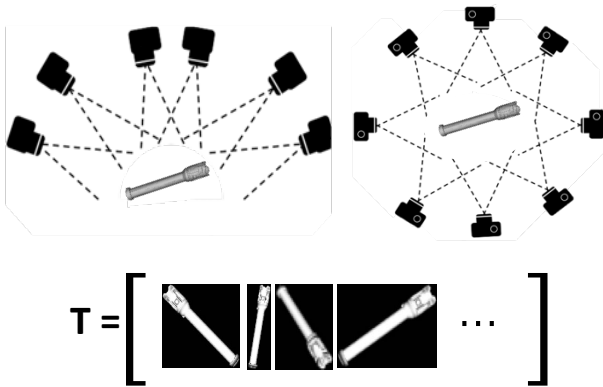
Figure 2. Sample set of templates $\mathcal{T}$ generated using different viewpoints sampled around the sample-tube.

The two localizers considered in this paper are a template-based object detector [38], and a region-proposal network [29]. Template-based object detection has more industrial maturity, and it shares building blocks with methods that ran on spacecrafts [47], [48]. Template matching lends itself to introspection and ad-hoc treatment of edge cases. By contrast, convolutional networks have almost no flight heritage, and their design is harder to validate, but their performance is often substantially better than template matching.

4. TEMPLATE-BASED OBJECT DETECTION

Our approach is based on Line2D [38], [39], an efficient template-matching based method that exploits color images to capture the appearance of the object in a set of templates covering different views. Because the viewpoint of each template is known, it also provides a coarse estimate of the 6D pose of the object when it is detected. In the remainder of this section, we give an overview of the Line2D method and how we use it to detect and localize sample-tubes.

Generating Templates using a 3D Model

Given a 3D CAD model of the sample-tube, a library of 2D templates are automatically generated that covers a full view hemisphere by regularly sampling viewpoints of the 3D model, as illustrated in Figure 2. During the generation of template library, it is important to balance the trade-off between the coverage of the object for reliability and the number of templates for efficiency. This is solved by recursively dividing an icosahedron, the largest convex regular polyhedron. The vertices of the resulting polyhedron give us the two out-of-plane rotation angles for the samples pose with respect to the coordinate center. In our experiments, two adjacent vertices are set to be approximately 10 degrees apart. In addition to the these two out of plane rotations, templates for different in-plane rotations are also created. Furthermore, templates at different scales are generated by using different sized polyhedrons, using a step size of 10 cm.

Gradient Orientation Features

For each sampled pose generated by the method described above, the sample-tube's silhouette is first computed by projecting its 3D model under this pose. The silhouette contour is then quickly obtained by subtracting the eroded silhouette from its original version. Next, all the color gradients that lie on the silhouette contour are computed. Image gradients are chosen as features because they have been proved to be

more discriminant than other forms of representation [49]. Additionally, image gradients are often the only reliable image cue when it comes to texture-less objects. Furthermore, considering only the orientation of the gradients and not their norms makes the measure robust to contrast changes, and taking the absolute value of cosine between them allows it to correctly handle object occluding boundaries: It will not be affected if the object is over a dark background, or a bright background.

Template Matching

Template matching is done by measuring the similarity between an input image $\mathcal{I}$, and a reference image $\mathcal{O}$ of the sample-tube centered on a location c in the image $\mathcal{I}$ by comparing the gradient orientation features. A model or template $\mathcal{T}$ is defined as a pair $\mathcal{T} = (\mathcal{O}, \mathcal{P})$, where $\mathcal{P}$ specifies a region in $\mathcal{O}$. The template can then be compared with a region at location c in a test image $\mathcal{I}$ based on a modified version of the similarity measure proposed by Steger [50]:

$$\varepsilon = \sum_{r \in \mathcal{P}} \Big(\max_{t \in \mathcal{R}(c+r)} \big| \cos \big(\mathrm{ori}(\mathcal{O}, r) - \mathrm{ori}(\mathcal{I}, t) \big) \big| \Big) \quad (1)$$

where $\mathrm{ori}(\mathcal{O}, r)$ is the gradient orientation at location r and $\mathcal{R}(c + r)$ defines the neighbourhood of size T centered on location $c + r$.

In order to avoid evaluating the $\max$ operator in Equation 1 every time a new template must be evaluated against an image location, a binary representation of the feature space is used [51]. First, the gradient orientation map is quantized by dividing it into n_0 equal spacing. To make the quantization robust to noise, we assign to each location the gradient whose quantized orientation occurs most often in a 3x3 neighborhood. Next, the possible combination of orientation spreads to a given location m is encoded using a binary string. These strings are then used as indices to access lookup tables for fast pre-computation of the similarity measure. Since the lookup tables are computed offline and shared between the templates, matching several templates against the input image can be done very fast once the maps are computed.

5. DATA-DRIVEN SEGMENTATION

Region Proposal CNNs

Modern CNNs are powerful models that leverage large labeled datasets in order to automatically derive visual feature hierarchies directly from the data such that a learning task, e.g. object classification, can be solved. In this section we describe our use of the popular CNN-based object instance segmentation model known as Mask R-CNN [29] for tube localization. The "R" in "R-CNN" stands for "region-based" and refers to a class of CNN networks that typically extend the classification capabilities of their forebears (e.g. [52], [53], [54]) by adding a regression head to the architecture such that they can predict continuous values for object region proposal as well as object class labels. This can be used to predict bounding box parameters of the objects under consideration, for example, such that the objects may be localized within images as opposed to solely having their presence detected [55].

Instance Segmentation with Mask R-CNN

Mask R-CNN builds on its predecessors Fast R-CNN [26] and Faster R-CNN [27], which were limited to bounding

3

box prediction, by additionally allowing for segmentation masks to be predicted. While related models like the Fully Convolutional Network (FCN) [28] have dealt with *semantic segmentation*, that is segmenting an image and associating the segments to the object classes to which they belong, Mask R-CNN tackles the more challenging problem of *instance segmentation* whereby individual instances of objects from those classes must be identified and segmented. For our purposes in solving the task of sample tube localization, all of these methods are viable candidates, but the instance segmentation provided by Mask-RCNN enables candidate tube instances to be directly identified and segmented without needing to further process the bounding box or segmentation results provided by the other methods.

Mask-RCNN uses the architecture of Faster R-CNN to both predict the class label and to regress the bounding box parameters, but it augments this with an FCN-based mask prediction branch in order to additionally predict the regional segmentation masks in parallel. There is some generality to the architecture in that a number of possible CNN backbones can be used for the convolutional feature detection component of the Faster R-CNN branch, e.g. ResNet [52], ResNeXt [53] or Feature Pyramid Networks (FPN) [54]. In our experiments for this work we used ResNet-50 exclusively [52]. For training, Mask R-CNN makes use of a multi-task loss $L = L_{cls} + L_{box} + L_{mask}$ that is comprised of the classification loss L_{cls} and the bounding-box loss L_{box} from [26], as well as a segmentation mask loss L_{mask}. L_{mask} is defined by taking the average binary cross-entropy loss over the per-pixel sigmoid-activated output of the mask prediction branch which is made up of K $m \times m$ resolution output masks for each predicted region of interest, where K is the number of ground-truth classes.

Using Mask R-CNN for Tube Localization

As is typical, the Mask R-CNN model we employ uses model weights pre-trained on the ImageNet dataset [25], [56] and transfer learning is used to fine-tune the weights for our particular task. The tube localization problem presented here is relatively simple in terms of the classification component since we need only detect the presence or absence of tubes in the scene, so we make use of only a single *"tube"* class label alongside the background class. With regard to localizing the tube, given the fetch rover grasping requirements, as well as for a direct comparison with the template matching method described in the previous section, we are less interested in the bounding box output than we are in obtaining the segmentation masks. Thus, although we annotate the training data with each of the class, bounding box and segmentation mask ground truth labels and train our Mask R-CNN model with all three of them, for the purposes of this paper, at inference time we discard the bounding box output and we restrict our subsequent experimental analyses to the segmentation masks.

6. BENCHMARKING DATASET

Camera Acquisition Setup

We constructed a camera acquisition setup made up of four FLIR BlackFly S cameras (5472×3648, color, $77°$ field of view) that form two stereo pairs with baselines of 20cm and 40cm. The cameras, optics and the overall acquisition setup (baselines and heights) are representative of the Perseverance rover's onboard cameras - EECAMs [57]. During image acquisition, we set the camera tripod at two different heights

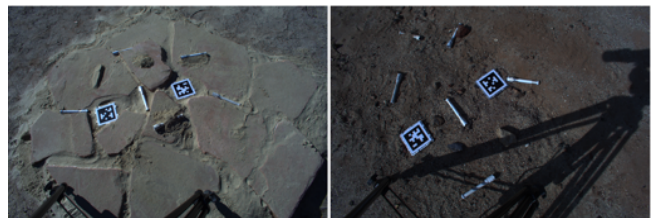(a) Camera setup at the Mars Yard at JPL.
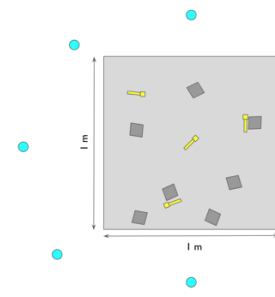
(b) Camera positioning around tubes.

(c) Sample images on flagstone (left) and CFA 6 rocks (right).

(d) Sample images on ditch (left) and riverbed (right).

(e) Sample images with tag-mounted tube: flagstone (left) and CFA 2 (right).

Figure 3. Outdoor dataset: (a) testbed, (b) capture grid, (c-d) sample images.

of 1m and 2m to simulate images acquired by HazCams and NavCams, respectively. Also, the cameras were covered by an aluminum plate serving as heat shield for extended use under sunlight (as shown in Fig 3a).

Data Collection

To evaluate the performance and robustness of our methods, we captured a dataset of outdoor images in JPL's Mars Yard. The curated set of images represents both nominal and adverse environmental conditions we expect SFR to face on Mars. An outdoor dataset provides us images of realistic scenes that contain: 1) diverse and varying terrain, informed by discussions with Mars geologists who plan where sample tubes could be dropped; 2) natural shadows and lighting that create appearance variation for the tube we are detecting. Additionally, we further vary the tube appearance by considering another adverse condition, object occlusion. Specifically, we achieve this by positioning tubes next to rocks or terrain features to induce varying levels of occlusion, and also covering

tubes with dust. We vary the amount of dust coverage in a few ways: 1) sprinkling a light layer of dust; 2) building up a "dune" on the side of a tube; 3) totally covering a partial section of the tube with a mound of sand.

Ground Truth Annotations—Finally, we enable quantitative evaluation of detection results against the dataset by providing ground truth segmentation masks with associated bounding boxes for 2D object detection benchmarking and 6DOF poses for a subset of the data, for pose estimation evaluation. The segmentation masks are manually annotated using the coco-annotator tool [58]. Ground truth 6D poses are obtained by rigidly mounting a pair of AprilTags [59] to one of the tubes. Two tags (see Fig-3e) were chosen for robustness since the occurrence of shadows sometimes hindered the tag detection algorithm. The tags are also mounted a sufficient distance away from the tube to not interfere with detection. During construction of the scene, care was also taken to cover the mount with sand so the tube appears as an isolated object.

The dataset comprises the following:

• One subset of images that does not contain the ground-truth tube with the mounted AprilTags (*no-tags* dataset). This set of images contains four terrain types: two terrain types constructed with consultation from the Mars geologists, and two extra ones we created based on other interesting features we saw in the Mars Yard to add more scene diversity (see Figs. 3c and 3d):
  – "Flagstone": broken stone slabs covered with a thin layer of dust emulating fractured bedrock on Mars.

  – "CFA6": a rock distribution of *cumulative fractional area* (CFA – a measure of rock density) equal to 6%. These are the smallest rocks that are still visible from orbit to guide the choice of depot location. Rocks encountered in practice would only be this big or smaller.

  – "Ditch", and "Riverbed": These 2 terrain types were simply named for the varying levels of surface depression we observed in the terrain.
• Another of subset of images that contains the tube mounted with AprilTags (*with-tags* dataset). This set of images contains the three terrain types we expect to experience on Mars:
  – "Flagstone" and "CFA6": same as in the other set.

  – "CFA2": small pebbles that are only visible from rover surface imagery and not orbit.
• Capture conditions common to both sets of images:
  – images taken in one of two different capture times: "am" (10am to 12pm), or "pm" (3pm to 5pm)

  – 5 or 6 sample tubes with variable visibility: unoccluded, or partially occluded by rocks and/or sand and dust.

  – images taken in one of 8 camera tripod positions on a circle around the scene, see Fig. 3b)

  – 2 camera heights at each stop (1m and 2m)

  – 4 cameras: 2 stereo pairs of baseline 20 and 40 cm

The dataset in total contains 824 images (256 with the tag-mounted tube and 568 without), and 4852 annotated instances of tubes, out of which 256 of those have an associated 6D pose. These images were collected over 2 separate days.

7. EXPERIMENTS AND RESULTS

In this section, we present experiments for both quantitative and qualitative performance analysis of the two object localizers for sample-tube localization on a Mars-like en-
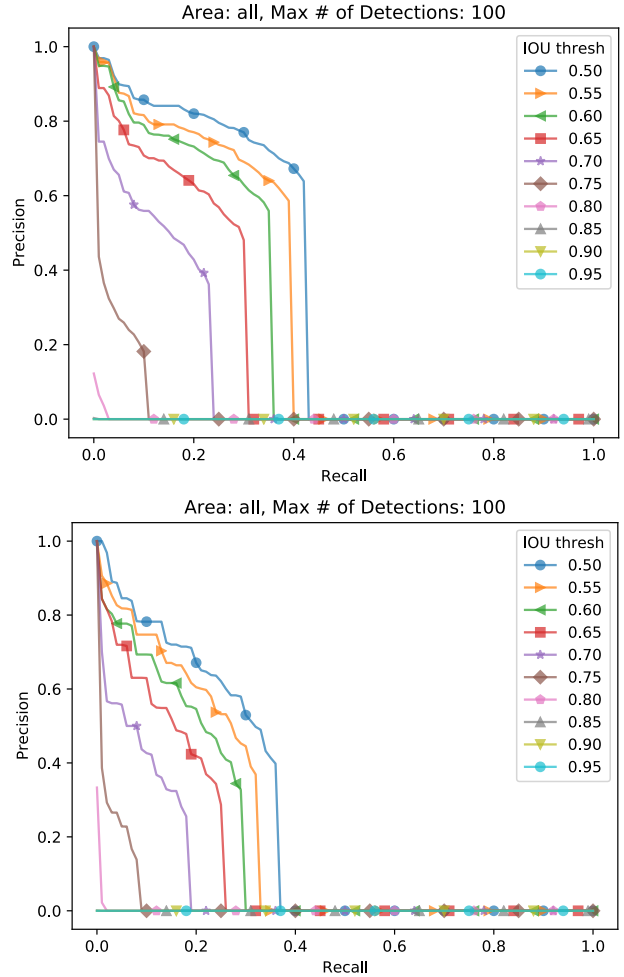

Figure 4. Precision Recall curves for the (a) no-tags and (b) with-tags datasets using the Line2D object detector. [5pt]

Dataset	Method	AP [.5]	AR [.5:.05:.95]
no-tags	Line2D	0.345	0.184
with-tags	Line2D	0.255	0.153

Table 1. Quantitative results using the Line2D object detector. AP and AR IoU thresholds are shown in square brackets.

vironment. All the experiments were conducted using the benchmark dataset described in Section 6.

Evaluation Metrics

In order to evaluate and compare the template matching and data-driven segmentation methods presented here, we make use of some of the de facto standard statistical metrics popularized in the literature and in computer vision contests, namely average precision (AP) and average recall (AR) values and precision-recall (PR) curves.

Average Precision—Various interpretations of the AP metric have been proposed over the years, particularly in the
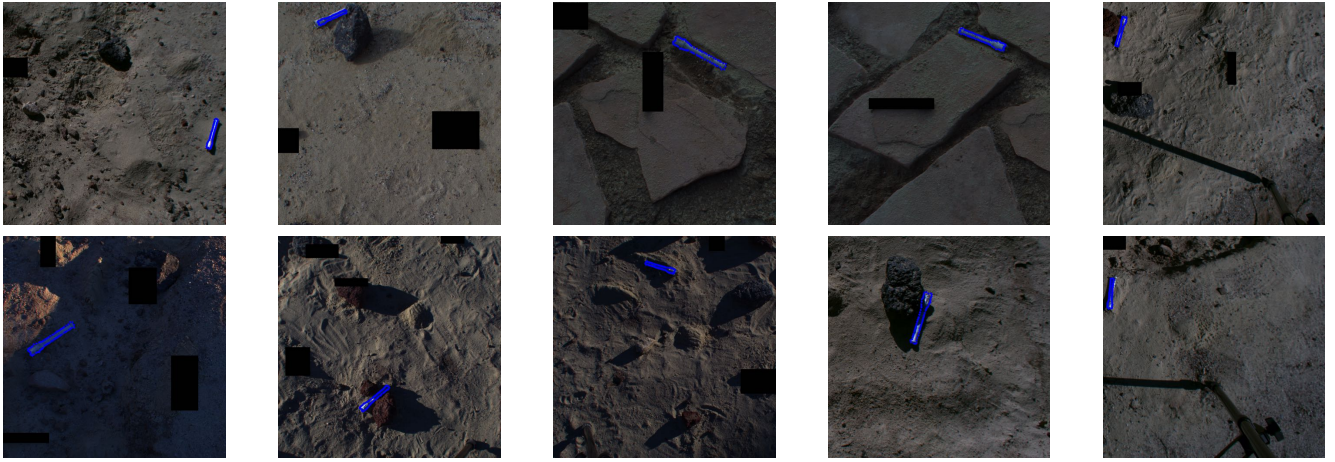
5

Figure 5. Example true positive results for sample-tube detection using the Line2D detector.
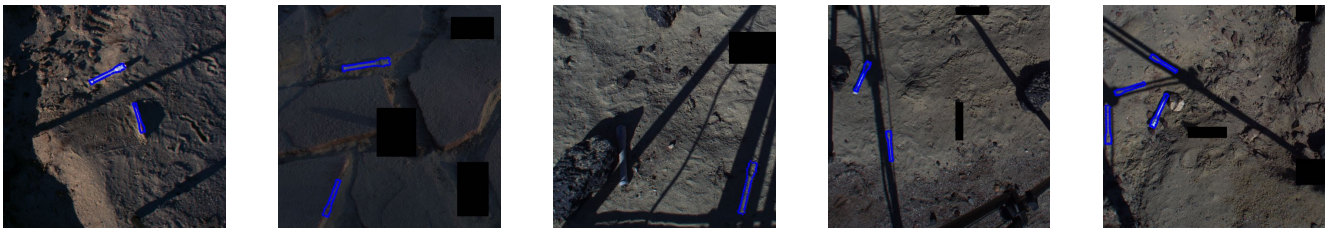

Figure 6. Examples of failure cases for sample-tube detection using the the Line2D detector.

PASCAL VOC [60] and MS COCO [61] challenges. In all cases, given an intersection-over-union (IoU) threshold for determining whether a prediction should be scored as a true positive (TP) or a false positive (FP) based on the amount of overlap between the predicted mask and the ground truth mask, cumulative precision and recall values are collated in a confidence-score-ranked PR table. These values may then either be used to produce PR curves or to calculate the AP. The variation between the PASCAL VOC and MS COCO metrics lies in how the AP is calculated using these values.

The PASCAL VOC 2008 AP metric divided the recall dimension into 11 points, interpolated the precision values and averaged them, whereas PASCAL VOC 2010-2012 estimated the area under the curve. The MS COCO metric by comparison, which we use for our evaluations here, takes a 101-point AP interpolation while also evaluating the AP across multiple IoU threshold values and/or averaging over a range of thresholds. Unless otherwise stated, in the analyses below, we use the MS COCO AP at IoU $= 0.5$ with all area sizes and maximum 100 detections, and we use the MS COCO AR averaged over a range of 10 IoU thresholds between 0.5 and 0.95 with a step size of 0.05 with the same area and max detection criteria. With regard to the mission goals, since we plan to use a 6D pose estimator to refine the tube localization post-detection, we can afford to use a looser AP overlap criterion so long as precision is maximized: we would rather avoid falsely labeling tubes on the first pass than highly accurately segmenting those tubes that we do correctly detect.

Precision-Recall Curves—In the figures presented below, PR curves are shown where the interpolated precision values are used for the 101 recall points used by the MS COCO AP metric. Plots are shown for PR curves for a range of 10 IoU thresholds between 0.5 and 0.95 with a step size of 0.05.

Performance Analysis of Template Matching

The Line2D detector was trained using 7000 templates generated from the 3D model of the sample-tube. The test data consisted of images from the benchmark dataset, down-scaled to half-resolution and randomly cropped to 1024x1024. The reasons to generate cropped images were two-fold. First, to match the input dimensions of the learning-based localizer for a fair comparison. Second, larger image size also increases the computational complexity of the algorithm. Table 1 shows the quantitative performance of the template matching based object localizer on the benchmark dataset using the AP and AR metrics. Furthermore, we also plot the PR curve in Figure 4 to demonstrate the inherent tradeoff between the precision and recall of our detector as a function of different threshold values. The overall quantitative performance of the template-based matching is encouraging, especially given the challenging and adversarial scenarios presented in our dataset. The low AP and AR values can be explained by the large number of false positives and false negatives, typical of most model-based algorithms.

Next, we qualitatively analyze the the performance of the detector as a function of different terrain types and environmental conditions. Some example results from successful detections using the Line2D detector are presented in Figure 5. It can be observed that the template matching algorithm is able to robustly detect sample-tubes in a wide variety of terrain (flat ground vs. cracks of bedrock) and illumination conditions (well-lit vs partial-shadow vs completely in shadow). The shape-based similarity measure also allows the template matching to be robust to partial occlusions (see examples where tubes are places next to a rock).

Finally we also look at some of the systematic failure cases where the template matching based detector's performance
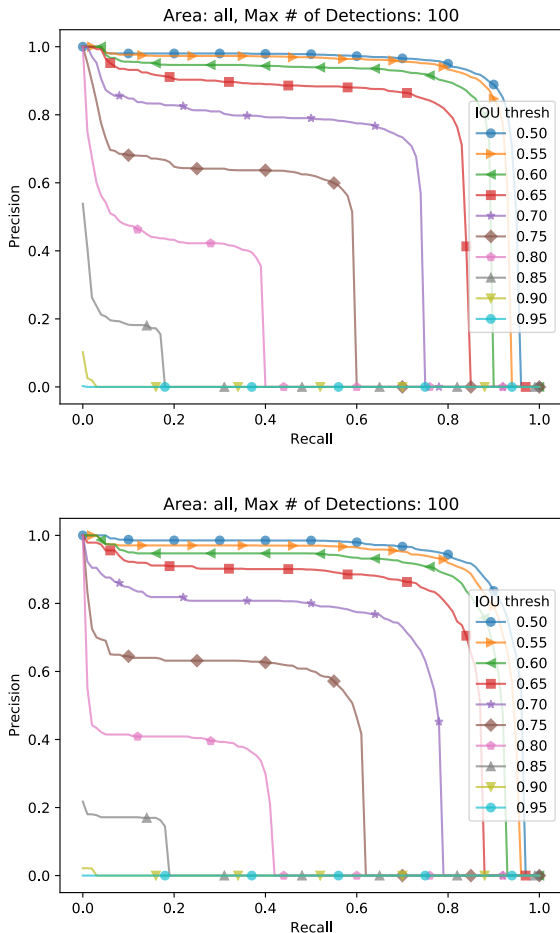
Figure 7. Precision-recall curves for the (a) no-tags and (b) with-tags datasets using the Mask R-CNN object detector.

Train	Test	AP [.5]	AR [.5:.05:.95]
with-tags	no-tags	0.911	0.555
no-tags	with-tags	0.918	0.575

Table 2. Quantitative results using the Mask R-CNN object detector. AP and AR IoU thresholds are shown in square brackets.

degrades. The most notable failure mode that degrades the performance of template matching was observed in the presence of other entities in the scene that have similar contour features compared to the sample-tubes. Two specific examples of this that can be observed in Figure 6, where object shadows and bedrocks - both have long edges with similar feature response to that of the long edges of the sample-tube. Another failure mode is related to occlusions. Template matching based algorithms perform poorly in the presence of significant occlusions - either from dust coverage or object occlusions.

Performance Analysis of Data-driven Segmentation

Two Mask R-CNN models were trained. One was trained on the with-tags dataset and evaluated on the no-tags dataset,

while the other was trained on the no-tags dataset and evaluated on the with-tags dataset. In each case the training data consists of 1200×1200 images that are converted to grayscale, box-filtered and randomly cropped to 1024×1024 to match the input size expected by the network and to allow for data augmentation. Our decision to use grayscale images reflects the intuition that color is less important for this task than edge and texture features and that the network might struggle to learn color invariance. In addition, the likelihood that the tubes would be coated with a thin layer of dust on Mars would mean that the tubes may acquire a more yellow hue than their original grey coloring that the network would be trained on. As for the box filtering, given the limited quantity of training data available, it may be helpful to prime the input to de-emphasize intensity over edge features. The test data consisted of 1024×1024 images, also converted to grayscale and box filtred, in which a whole single tube is present within the image boundaries. When the no-tags data was used for training, the with-tags data was used for testing, and vice versa.

Quantitative AP and AR results are presented in Table 2, PR curves are presented in Figure 7, and qualitative results samples are presented in Figures 8 and 9. As was expected, the data-driven learning approach of Mask R-CNN provides a substantial average increase in performance over the template-matching method both in terms of precision and recall. The PR curves also demonstrate that if we can tolerate detection overlap with ground-truth in the $0.5 - 0.6$ IoU threshold range, we can expect reasonably optimal precision-recall. The results samples in Figures 8 and 9 show that Mask R-CNN can cope well even in difficult situations where object shadows overlap the tubes. Conversely, it can also produce false positives in cases where there is not enough overlap between the detection and the ground-truth or sometimes fail completely in cases where features from non-tube objects seem tube-like.

Performance Analysis of 6D Pose Estimation

Using the *with-tags* dataset with ground-truth 6D pose information, we can evaluate the pose estimation accuracy of the Line2D detector. Figure 10 shows histograms of the orientation and translation error magnitudes. Since the detector does not know which is the tag-mounted tube, we only consider detections with IOU > 0.5 for evaluation of the 6D pose accuracy. If there is indeed a detection in the neighborhood, there should only be one since additional ones will have been removed by non-max suppression. Moreover, we consider the orientation error in a couple of ways. First, we focus on in-plane rotation and report error along the main axis of the tube since this is the most relevant for the downstream task of tube pickup. Also, from the histograms, we see that there are a few outliers with large orientation errors. These are likely detections where the matched template is "flipped" $180°$ due to the most prominent object features being the two parallel edges that define the body of the tube. Ignoring these instances allows us to observe the nominal orientation errors. Furthermore, in practice an orientation flip should not affect ability for tube pickup.

8. Conclusion and Future Work

Autonomous localization and retrieval of sample-tubes for the Mars Sample Return mission is a challenging task, but one that is necessarily to accommodate the mission timeline for the Sample Fetch Rover. In this work, we studied two machine-vision based approaches to autonomously
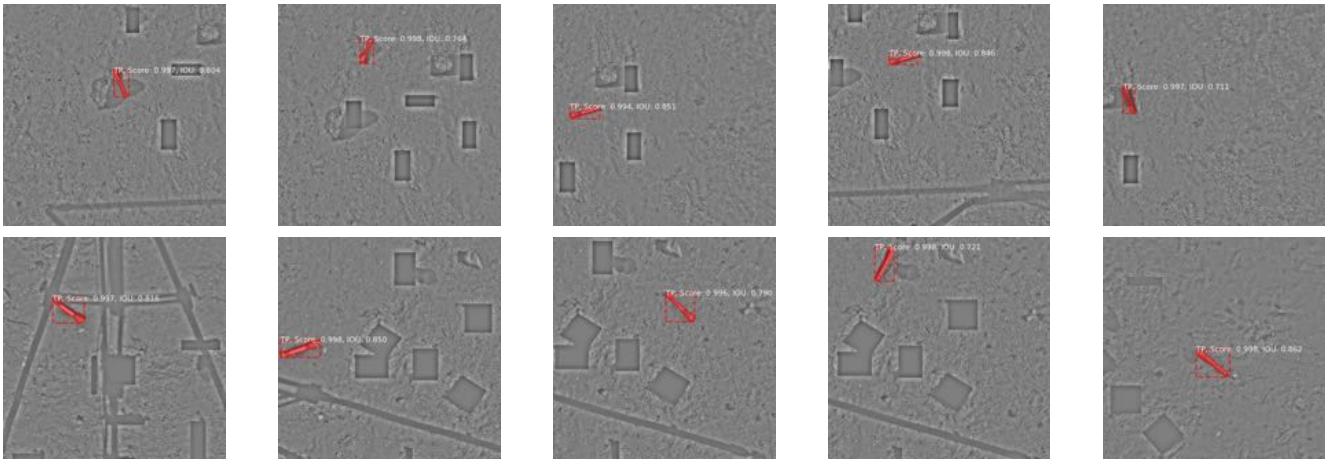
Figure 8. Example true positive results for sample-tube detection using the Mask R-CNN detector. The top row shows results from the no-tags test set; the bottom row shows results from the with-tags test set.

Figure 9. Example false positive cases for the Mask R-CNN detector. The first three images are from the no-tags test set; the last two images are from the with-tags test set.
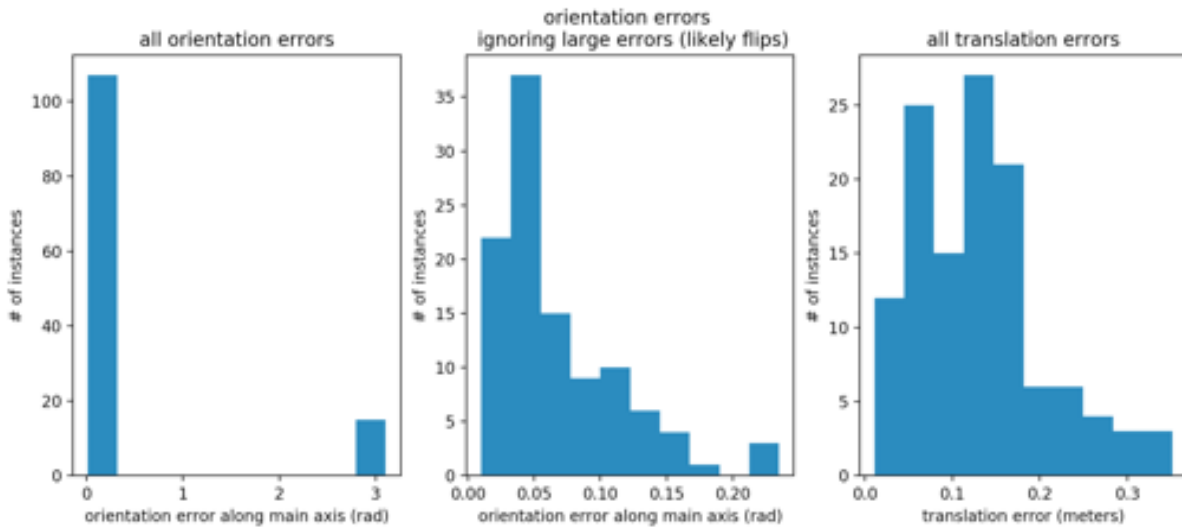

Figure 10. 6D pose estimation results using Line2D

detect and localize sample-tubes. Our top-level goal was to understand the trade-off between the different classes of algorithms: model-based and data-driven. Towards this end, we also collected a large benchmark dataset of sample-tube images, in a representative outdoor environment and annotated it with ground truth segmentation masks and locations for performance analysis. In summary, both methods have complimentary advantages. While, learning-based methods are considerably superior in terms of performance, they are fundamentally black-box from a design perspective. This presents a major challenge for Verification and Validation (V&V) and difficulty in flight infusion. On the other hand, classical methods such as Template matching do not match the performance of their learning based counterparts, but are easier to design, implement and V&V. In future work, we plan to validate the performance of autonomous tube-pickup through end-to-end demonstration on an analog Fetch rover.

References

[1] S. W. Squyres, R. E. Arvidson, D. L. Blaney, B. C. Clark, L. Crumpler, W. H. Farrand, S. Gorevan, K. E. Herkenhoff, J. Hurowitz, A. Kusack *et al.*, "Rocks of the columbia hills," *Journal of Geophysical Research: Planets*, vol. 111, no. E2, 2006.

[2] J. P. Grotzinger, "Analysis of surface materials by the curiosity mars rover: Introduction," *Science*, vol. 341, no. 6153, pp. 1475–1475, 2013.

[3] S. S. Board, N. R. Council *et al.*, *Vision and voyages for planetary science in the decade 2013-2022*. National Academies Press, 2012.

[4] B. K. Muirhead, A. K. Nicholas, J. Umland, O. Sutherland, and S. Vijendran, "Mars sample return mission concept status," *Acta Astronautica*, 2020.

[5] B. Muirhead, C. Edwards, A. Eremenko, A. Nicholas, A. Farrington, A. Jackman, S. Vijendran, L. Duvet, F. Beyer, and S. Aziz, "Sample retrieval lander concept for a potential mars sample return campaign," *LPICo*, vol. 2089, p. 6369, 2019.

[6] M. N. D. S. A. Group, "Science priorities for mars sample return," 2008.

[7] S. McLennan, M. Sephton, D. Beaty, M. Hecht, B. Pepin, I. Leya, J. Jones, B. Weiss, M. Race, J. Rummel *et al.*, "Planning for mars returned sample science: final report of the msr end-to-end international science analysis group (e2e-isag)," *Astrobiology*, vol. 12, no. 3, pp. 175–230, 2012.

[8] D. Beaty, M. Grady, H. McSween, E. Sefton-Nash, B. Carrier, F. Altieri, Y. Amelin, E. Ammannito, M. Anand, L. Benning *et al.*, "The potential science and engineering value of samples delivered to earth by mars sample return," *Meteoritics and Planetary Science*, vol. 54, no. 3, 2019.

[9] B. Sherwood, D. B. Smith, R. Greeley, W. Whittaker, G. R. Woodcock, G. Barton, D. W. Pearson, and W. Siegfried, "Mars sample return: Architecture and mission design," in *Proceedings, IEEE Aerospace Conference*, vol. 2. IEEE, 2002, pp. 2–536.

[10] R. Mattingly, S. Matousek, and F. Jordan, "Continuing evolution of mars sample return," in *2004 IEEE Aerospace Conference Proceedings (IEEE Cat. No. 04TH8720)*, vol. 1. IEEE, 2004.

[11] R. Mattingly and L. May, "Mars sample return as a campaign," in *IEEE Aerospace Conference*, 2011.

[12] B. K. Muirhead, A. K. Nicholas, J. Umland, O. Sutherland, and S. Vijendran, "Mars sample return campaign concept status," *Acta Astronautica*, vol. 176, pp. 131–138, 2020.

[13] R. Volpe, E. Baumgartner, P. Scheaker, and S. Hayati, "Technology development and testing for enhanced mars rover sample return operations," in *2000 IEEE Aerospace Conference. Proceedings (Cat. No. 00TH8484)*, vol. 7. IEEE, 2000, pp. 247–257.

[14] R. Weisbin, G. Rodriguez, S. Schenker, H. Das, S. Hayati, T. Baumgartner, M. Maimone, I. Nesnas, and A. Volpe, "Autonomous rover technology for mars sample return," in *Artificial Intelligence, Robotics and Automation in Space*, vol. 440, 1999, p. 1.

[15] G. R. Osinski, M. Battler, C. M. Caudill, R. Francis, T. Haltigin, V. J. Hipkin, M. Kerrigan, E. A. Pilles, A. Pontefract, L. L. Tornabene *et al.*, "The canmars mars sample return analogue mission," *Planetary and Space Science*, vol. 166, pp. 110–130, 2019.

[16] P. Younse, C. Y. Chiu, J. Cameron, M. Dolci, E. Elliot, A. Ishigo, D. Kogan, E. Marteau, J. Mayo, J. Munger *et al.*, "Concept for an on-orbit capture and orient module for potential mars sample return," in *2020 IEEE Aerospace Conference*. IEEE, 2020, pp. 1–22.

[17] S. Perino, D. Cooper, D. Rosing, L. Giersch, Z. Ousnamer, V. Jamnejad, C. Spurgers, M. Redmond, M. Lobbia, T. Komarek *et al.*, "The evolution of an orbiting sample container for potential mars sample return," in *2017 IEEE Aerospace Conference*. IEEE, 2017, pp. 1–16.

[18] K. Edelberg, J. Reid, R. McCormick, L. DuCharme, E. Kulczycki, and P. Backes, "Autonomous localization and acquisition of a sample tube for mars sample return," in *AIAA SPACE 2015 Conference and Exposition*, 2015, p. 4483.

[19] J. Papon, R. Detry, P. Vieira, S. Brooks, T. Srinivasan, A. Peterson, and E. Kulczycki, "Martian fetch: Finding and retrieving sample-tubes on the surface of mars," in *IEEE Aerospace Conference*, 2017.

[20] B. Lee, R. Detry, J. Moreno, D. D. Lee, and E. Kulczycki, "Monocular visual pose estimation via online sampling for mars sample-tube pickup," in *2018 IEEE Aerospace Conference*. IEEE, 2018, pp. 1–8.

[21] T.-H. Pham, W. Seto, S. Daftry, A. Brinkman, J. Mayo, Y. Cheng, C. Padgett, E. Kulczycki, and R. Detry, "Rover localization for tube pickup: Dataset, methods and validation for mars sample return planning," in *2020 IEEE Aerospace Conference*. IEEE, 2020, pp. 1–11.

[22] Z. Zou, Z. Shi, Y. Guo, and J. Ye, "Object detection in 20 years: A survey," *arXiv preprint arXiv:1905.05055*, 2019.

[23] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the seventh IEEE international conference on computer vision*, vol. 2. Ieee, 1999, pp. 1150–1157.

[24] F. Rothganger, S. Lazebnik, C. Schmid, and J. Ponce, "3d object modeling and recognition using local affine-invariant image descriptors and multi-view spatial constraints," *International journal of computer vision*, vol. 66, no. 3, pp. 231–259, 2006.

[25] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.

[26] R. Girshick, "Fast R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1440–1448.

[27] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in *Advances in Neural Information*

Processing Systems 28, C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, Eds. Curran Associates, Inc., 2015, pp. 91–99.

[28] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3431–3440.

[29] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2961–2969.

[30] M. Ono, B. Rothrock, K. Otsu, S. Higa, Y. Iwashita, A. Didier, T. Islam, C. Laporte, V. Sun, K. Stack *et al.*, "Maars: Machine learning-based analytics for automated rover systems," in *2020 IEEE Aerospace Conference*. IEEE, 2020, pp. 1–17.

[31] N. Abcouwer, S. Daftry, S. Venkatraman, T. del Sesto, O. Toupet, R. Lanka, J. Song, Y. Yue, and M. Ono, "Machine learning based path planning for improved rover navigation (pre-print version)," in *2021 IEEE Aerospace Conference*. IEEE, 2021.

[32] M. S. Costa and L. G. Shapiro, "3d object recognition and pose with relational indexing," *Computer Vision and Image Understanding*, vol. 79, no. 3, pp. 364–407, 2000.

[33] P. David and D. DeMenthon, "Object recognition in high clutter images using line features," in *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, vol. 2. IEEE, 2005, pp. 1581–1588.

[34] I. Weiss and M. Ray, "Model-based recognition of 3d objects from single images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 116–128, 2001.

[35] C. M. Cyr and B. B. Kimia, "A similarity-based aspect-graph approach to 3d object recognition," *International Journal of Computer Vision*, vol. 57, no. 1, pp. 5–22, 2004.

[36] D. W. Eggert, K. W. Bowyer, C. R. Dyer, H. I. Christensen, and D. B. Goldgof, "The scale space aspect graph," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 11, pp. 1114–1130, 1993.

[37] M. Ulrich, C. Wiedemann, and C. Steger, "Combining scale-space and similarity-based aspect graphs for fast 3d object recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 10, pp. 1902–1914, 2011.

[38] S. Hinterstoisser, C. Cagniart, S. Ilic, P. Sturm, N. Navab, P. Fua, and V. Lepetit, "Gradient response maps for real-time detection of textureless objects," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 5, pp. 876–888, 2011.

[39] S. Hinterstoisser, V. Lepetit, S. Ilic, S. Holzer, G. Bradski, K. Konolige, and N. Navab, "Model based training, detection and pose estimation of texture-less 3d objects in heavily cluttered scenes," in *Asian conference on computer vision*. Springer, 2012, pp. 548–562.

[40] H. Cai, T. Werner, and J. Matas, "Fast detection of multiple textureless 3-d objects," in *International Conference on Computer Vision Systems*. Springer, 2013, pp. 103–112.

[41] F. Tombari, S. Salti, and L. Di Stefano, "Unique signatures of histograms for local surface description," in *European Conference on Computer Vision*, 2010, pp. 356–369.

[42] C.-Y. Tsai, C.-C. Yu *et al.*, "Real-time textureless object detection and recognition based on an edge-based hierarchical template matching algorithm," *Journal of Applied Science and Engineering*, vol. 21, no. 2, pp. 229–240, 2018.

[43] T.-T. Do, M. Cai, T. Pham, and I. Reid, "Deep-6dpose: Recovering 6d object pose from a single rgb image," *arXiv preprint arXiv:1802.10367*, 2018.

[44] W. Kehl, F. Manhardt, F. Tombari, S. Ilic, and N. Navab, "Ssd-6d: Making rgb-based 3d detection and 6d pose estimation great again," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 1521–1529.

[45] Y. Xiang, T. Schmidt, V. Narayanan, and D. Fox, "Posecnn: A convolutional neural network for 6d object pose estimation in cluttered scenes," *arXiv preprint arXiv:1711.00199*, 2017.

[46] T.-H. Pham, W. Seto, S. Daftry, B. Ridge, J. Hansen, T. Thrush, J. Mayo, Y. Cheng, C. Padgett, E. Kulczycki, and R. Detry, "Rover Navigation for Mars Sample Return Planning: Relocalization in Changing Environments by Virtual Template Synthesis and Matching," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, Xi'an, China, May 2021 (Under Review).

[47] R. Francis, T. Estlin, G. Doran, S. Johnstone, D. Gaines, V. Verma, M. Burl, J. Frydenvang, S. Montaño, R. Wiens *et al.*, "Aegis autonomous targeting for chemcam on mars science laboratory: Deployment and results of initial science team use," *Science Robotics*, vol. 2, no. 7, 2017.

[48] W. Kim, R. Steele, A. Ansar, K. Ali, and I. Nesnas, "Rover-based visual target tracking validation and mission infusion," in *Space 2005*, 2005.

[49] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[50] C. Steger, "Occlusion, clutter, and illumination invariant object recognition," *International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences*, vol. 34, no. 3/A, pp. 345–350, 2002.

[51] R. Rios-Cabrera and T. Tuytelaars, "Discriminatively trained templates for 3d object detection: A real time scalable approach," in *Proceedings of the IEEE international conference on computer vision*, 2013, pp. 2048–2055.

[52] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[53] S. Xie, R. Girshick, P. Dollar, Z. Tu, and K. He, "Aggregated Residual Transformations for Deep Neural Networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1492–1500.

[54] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature Pyramid Networks for Object Detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2117–2125.

[55] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich

Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," in *2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2014, pp. 580–587.

[56] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, Dec. 2015.

[57] J. Maki, C. McKinney, R. Sellar, D. Copley-Woods, D. Gruel, D. Nuding, M. Valvo, T. Goodsall, J. McGuire, and T. Litwin, "Enhanced engineering cameras (eecams) for the mars 2020 rover," *LPICo*, vol. 1980, p. 4132, 2016.

[58] J. Brooks, "COCO Annotator," https://github.com/jsbroks/coco-annotator/, 2019.

[59] E. Olson, "Apriltag: A robust and flexible visual fiducial system," *2011 IEEE International Conference on Robotics and Automation*, 2011.

[60] M. Everingham, L. V. Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal Visual Object Classes (VOC) Challenge," *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, Jun. 2010.

[61] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common Objects in Context," in *Computer Vision – ECCV 2014*, ser. Lecture Notes in Computer Science, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds.   Springer International Publishing, 2014, pp. 740–755.

BIOGRAPHY

Shreyansh Daftry *is a Robotics Technologist at NASA Jet Propulsion Laboratory, California Institute of Technology. He received his M.S. degree in Robotics from the Robotics Institute, Carnegie Mellon University in 2016, and his B.S. degree in Electronics and Communication Engineering in 2013. His research interest lies is at intersection of space technology and autonomous robotic systems, with an emphasis on machine learning applications to perception, planning and decision making. At JPL, he has worked on mission formulation for Mars Sample Return, and technology development for autonomous navigation of ground, airborne and subterranean robots.*

Barry Ridge *is a postdoctoral scholar at the NASA Jet Propulsion Laboratory (JPL), California Institute of Technology. He has previously held postdoctoral positions at the Advanced Telecommunications Research Institute International, Kyoto, Japan, and the Jožef Stefan Institute, Ljubljana, Slovenia. He received a B.Sc. in Computer Applications from Dublin City University, Ireland, in 2002 and went on to study pure mathematics at the University of St Andrews, Scotland, where he received an M.Phil. in 2006. After being awarded a Marie Curie Fellowship to pursue doctoral studies at the University of Ljubljana, Slovenia, with* a focus on robotic learning of object affordances, he defended his Ph.D. in 2014. His research interests include cognitive robotics, computer vision, and machine learning, and at JPL he is currently working on robot vision and simulation capabilities for the Mars Sample Return and InVADER missions.

William Seto *is a Robotics Technologist at NASA's Jet Propulsion Laboratory. He joined JPL in 2017 after receiving his M.S. in Robotic Systems Development from Carnegie Mellon's Robotics Institute. He develops software to enable autonomous capabilities in maritime and terrestrial environments. His outside interests include soccer and chicken tenders.*

Tu-Hoa Pham *is a Robotics Technologist at the NASA Jet Propulsion Laboratory, Caltech Institute of Technology, currently working on machine vision for Mars Sample Return. He holds a Diplôme d'Ingénieur in Aerospace Engineering from ISAE-SUPAERO (2013), an M.Sc. in Applied Mathematics from Université Paul Sabatier (2013) and a Ph.D. in Robotics from Université de Montpellier (2016), which he conducted at the CNRS-AIST Joint Robotics Laboratory on the topic of force sensing from vision. Prior to joining JPL in 2018, he spent two years as a research scientist at IBM Research Tokyo, where he worked on deep reinforcement learning for robot vision and manipulation in the real-world.*

Peter Ilhardt *received his B.A. in Earth and Planetary Sciences from Northwestern University in 2013 and M.S. in Geosciences from Penn State University in 2016. He is currently a data science consultant at Capgemini and M.C.S. student at the University of Illinois. His previous work focused on spectroscopic analysis and chemical imaging of complex biogeochemical systems at Pacific Northwest National Laboratory and Penn State University. Before starting his current role, Peter interned with the Perception Systems Group at NASA JPL working on machine vision for Mars Sample Return.*

Gerard Maggiolino received a B.S. in Math-Computer Science from the University of California, San Diego in 2020 and will be attending Carnegie Mellon University for a Masters in Computer Vision in 2021. He has previously interned at NASA JPL (Pasadena), Accel Robotics (San Diego), and Elementary Robotics (Los Angeles) working on primarily Computer Vision, Machine Learning, and Software Development projects.

Mark Van der Merwe is a Robotics PhD student at the University of Michigan, Ann Arbor. He received his B.Sc. degree in Computer Science from the University of Utah in 2020. He interned at NASA JPL during the Summer of 2020, where he worked with the Perception Systems team on tube and rover localization for Mars Sample Return. His current research interests lie at the intersections of perception and action for robot manipulation.

Alexander Brinkman received his M.S in Robotic Systems Development from Carnegie Mellon's Robotic Institute, then joined the Robotic Manipulation and Sampling group at Jet Propulsion Laboratory in 2017. He develops manipulation software and autonomous capabilities to enable future sampling missions to Europa, Enceladus, Mars, and comets.

John Mayo is a robotics mechanical engineer in the Robotic Climbers and Grippers Group at JPL. John received a Bachelor of Science in Mechanical Engineering from Texas A&M in 2014 and Master of Science of the same from the Massachusetts Institute of Technology in 2016. As part of his graduate studies, John worked on hardware for the HERMES humanoid robot, developing a hybrid hand-foot device under direction of Sangbae Kim. Additionally, John co-founded and led the MIT Hyperloop Team to design and build a magnetically levitated vehicle and participated as a mentor in the new student-led shop, MIT Makerworks.

Eric Kulczycki 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 2020 Sample Caching Subsystem, and mechanical design of various mobility platforms. He has worked at JPL for over 15 years.

Renaud Detry is the group leader for the Perception Systems group at NASA's Jet Propulsion Laboratory (JPL). Detry earned his Master's and Ph.D. degrees in computer engineering and robot learning from ULiege in 2006 and 2010. He served as a postdoc at KTH and ULiege between 2011 and 2015. He joined the Robotics and Mobility Section at JPL in 2016. His research interests are perception and learning for manipulation, robot grasping, and mobility, for terrestrial and planetary applications. At JPL, Detry leads the machine-vision team of the Mars Sample Return surface mission, and he conducts research in autonomous robot manipulation and mobility for Mars, Europa, Enceladus, and terrestrial applications.