# Robust Vision-Based Pose Correction for a Robotic Manipulator using Active Markers

Lukas Meyer, Klaus H. Strobl, and Rudolph Triebel

**Preprint:** This is a preprint of the following chapter: Lukas Meyer et al., *Robust Vision-Based Pose Correction for a Robotic Manipulator Using Active Markers*, published in *Experimental Robotics*, edited by *Bruno Siciliano, Cecilia Laschi, and Oussama Khatib*, 2021, *Springer International Publishing*; reproduced with the permission of *Springer International Publishing*. The final authenticated version is available online at: http://dx.doi.org/10.1007/978-3-030-71151-1_47.

**Abstract** Robots with elastic or lightweight components are becoming common in research, but can suffer from undesired positioning imprecision, which motivates a vision-based pose correction of the manipulator. For robotic manipulators that operate outdoors and under changing illumination conditions, robustness of the vision components is of principal concern. We propose a monocular manipulator pose correction based on active markers which are detected by convergence criteria on the image gradient field. We show the capabilities of the method in several outdoor and indoor experiments, considering the use case of a planetary exploration rover prototype equipped with a lightweight robotic arm. The vision-based manipulator pose correction method proves to be successful despite back light, reflections, and image overexposure and additionally allows continued robot operation in the case of extrinsic camera decalibration.

**Fig. 1** Two operation scenarios of the LRU during the 2017 Etna Moon Analogue Campaign [10]: Collection of scientific equipment from a lander (left) and deployment of a geophone in the ground (right). Credit: DLR.

## 1 Motivation and Related Work

In 2017, the Lightweight Rover Unit (LRU) of the DLR performed several scientific experiments on the volcano Mt. Etna in Italy during a Moon Analogue campaign [10], as illustrated in Fig. 1.

The successful mission also revealed inaccuracies during manipulation and sensitivity of the system to extrinsic camera decalibration due to the robot's lightweight structure. We develop a vision-based pose correction (VBPC) method for manipulators to overcome these issues with the emphasis on the robustness for robot operations in challenging illumination conditions.

The LRU is scheduled to participate in a second Moon-analog scientific mission on the volcano Mt. Etna in June and July 2021 [15], a sequel to the 2017 campaign. We consider the LRU and the Etna mission scenario as our primary use case for VBPC, addressing three aspects: First, a correction of the manipulator pose is needed, which is applied after having finished a coarse approach towards a target and before starting the dexterous manipulation process. Second, the detected position of the markers within the camera image is used as initialization of a tracking approach with focus on robust visual servoing, which we currently develop. Lastly, infrequent comparison of the end effector pose between VBPC and forward kinematics aims at detecting an hand-to-eye decalibration.

In literature, several approaches exist for pose estimation of robotic manipulators using vision. A widely used technique is fitting the three dimensional robot model to available depth information [4, 14, 7]. Monocular tracking based on rendering CAD models is done by e.g. [5, 3, 13]. Machine learning based methods are found in [9, 11]. The mentioned methods require either dense depth information or extensive computation resources; neither of which is available to us.

Lukas Meyer, Klaus H. Strobl, Rudolph Triebel
German Aerospace Center (DLR), Institute of Robotics and Mechatronics, Wessling, Germany, e-mail: {firstname.lastname}@dlr.de

We therefore follow an active marker based approach that is more robust to illumination changes than passive markers. Likewise, [1, 2] use active markers for pose estimation. However, they either require that the markers are the brightest object in the image or require specific wavelength filters – requirements that are not available to us. We therefore follow a novel approach by detecting the convergence of the image gradients at the marker locations, which turns out to be robust w.r.t. most of the external disturbances.

The LRU (see Fig. 1) has a robotic arm mounted on its rear. It also features a stereo camera pan-tilt head for navigation and two fixed rear monocular cameras for close-up observation of upper and the lower sections of the manipulator's workspace. As end effector, the arm is equipped with the cylindrical *Envicon* docking interface [10] that has a ring of 12 light-emitting diodes (LEDs) which we use as active markers for VBPC. In this work, we consider the principal navigation camera and the upwards facing rear camera. The cameras operate at 14 Hz and 2 Hz respectively. The frequencies of the cameras are determined by other system components, therefore limiting the rear camera to updates of static poses, whereas the navigation camera allows for a continuous pose correction during manipulator movements.

## 2 Technical Approach

Our approach to correct the manipulator's pose is based on combining a-priori information from the forward kinematics with a marker detection based on the structure of the image gradient field and a subsequent fitting of the marker model.

### 2.1 Marker Detection

The common optical characteristic of our active markers – which is invariant under a high variability in illumination – is the convergence of the image gradient vectors towards the center of each marker as illustrated in Fig. 2. We calculate the *convergence score* of each pixel in a region of interest (ROI) by applying a modified *IRIS-filter* [8] to the gradient field.

The modified IRIS-filter calculates a convergence score $C$ for a pixel $(u, v)$ by considering the orientation of the nearby gradient vectors and additionally includes the magnitude of the gradient vectors to focus only on relatively bright regions. We initialize each pose correction step using the forward kinematics to define a ROI in the camera image and to obtain a prediction for the marker positions within the ROI. For each pixel, we define $N$ half-lines over the circular arc with a maximum radius $R_{\max}$. Each half-line has an associated unit-orientation-vector $\boldsymbol{r}_i$ with $i = 1, ..., N$. Along the $i$th half line, starting with a distance $R_{\min}$ from the center, we sample the gradient vectors $\boldsymbol{g}$, such that

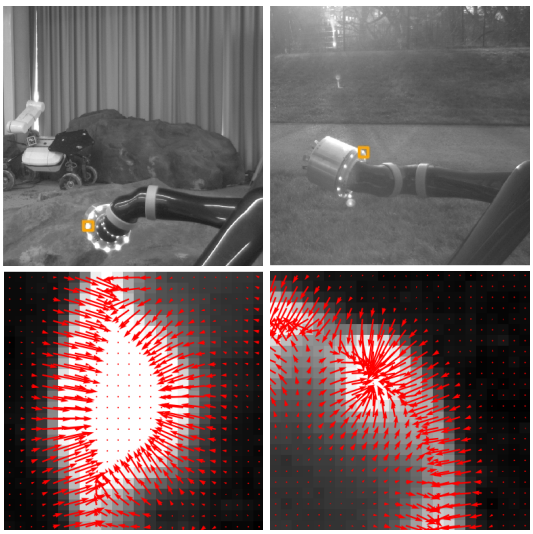

**Fig. 2** View of the manipulator from the navigation camera (top row), and detailed view of individual LEDs (orange box, shown in bottom row) with the respective gradient vector field (red). Different illumination conditions are shown: over exposure and reflections, indoors (left); and backlight, outdoors (right).

$$C_i(u,v) = \max_n \frac{1}{n} \sum_{m=R_{\min}}^{n} -\cos(\theta_m)\|\boldsymbol{g}_m\| \tag{1}$$

with $R_{\min} < n < R_{\max}$. The angle between $\boldsymbol{g}_m$ and the half-line is denoted by $\theta_m$. The overall score $C$ is the sum of all line-scores $C_i$. Marker candidates are the pixels with the highest values in the resulting IRIS-score image.

Finally, we compute the variance of the IRIS-score for each pixel as

$$\sigma^2(u,v) = \frac{1}{N} \sum_{i=0}^{N} (C_i(u,v) - \bar{C}(u,v))^2, \tag{2}$$

with $\bar{C}$ denoting the mean IRIS-score over all $N$ half-lines. The variance allows to penalize pixels that have a high IRIS-score along several half-lines, but a low score in others, e.g. pixels that correspond to circular segments or similar. The detected marker candidates are these pixels with the highest score after penalization.

## 2.2 Robust Pose Correction

The obtained marker candidates are fitted to the respective model of the LED-distribution. In our case, the markers are arranged as a circle, which allows for a random sample consensus (RANSAC) based fitting of an ellipse using the algorithm from [6]. To obtain the ellipse, theoretically, only five out of the 12 LEDs need to be visible. However, our experiments revealed that with five visible markers, the resulting ellipse becomes highly sensitive to noise in the marker position and is

therefore rejected by our a-priori informed outlier detection. Therefore, usually six markers need to be detected for a robust fitting, thus allowing for a 50% occlusion rate. Note that the approach can generally be used with any other type of marker pattern requiring only minor adaptation.

We compare the resulting ellipse with the ellipse computed from the a-priori prediction for outlier rejection. The position of the ellipse center is extracted to obtain the lateral location of the end effector w.r.t. the camera frame. The depth of the end effector position is calculated from the observed length of the ellipse major axis.

The initial result of the pose correction allows to obtain an image-to-world correspondence between the IRIS-maxima and the respective LEDs from the known model. This is used for a final refinement step using an iterative implementation of the perspective n-point (PnP) algorithm [12], with the position from the ellipse fitting and the orientation from the forward kinematics for initialization. The result from the PnP iteration provides the final pose of the LEDs w. r. t. the camera frame and is used for robot operations.

For our pose correction, we focus primarily on robustness towards the challenging and varying illumination conditions. We expect that the VBPC will experience increasingly false detections with increasing disturbances. For this purpose, we rely on a strict outlier rejection policy with the tradeoff of a decreased update rate of the pose. As outlier criteria, we mainly consider the fact that the kinematic pose is erroneous, however this error is bounded. We furthermore consider the shape of the ellipse w.r.t. the a-priori kinematic information. To account for a possible low update rate, we map the corrected end effector pose as offset to the end effector frame that is calculated by the forward kinematics. This allows to maintain a locally valid pose correction that only requires infrequent VBPC updates.

## 2.3 Concept Verification

The results of our algorithm are illustrated in Fig. 3. The circle of the LEDs is detected in all scenarios. The end effector frame from the VBPC (colored) is shown together with the a-priori prediction from the forward kinematics (gray). The first row (a to c) shows the VBPC results considering the principal navigation cameras. LED detection via the upward facing rear camera is shown in (d) to (f), with (f) being generally the most challenging scenario due to bright cluttered background. In all scenarios, VBPC allows for an improved pose estimate.

LED detection and ellipse fitting is successful for most of the frames. With increasing disturbances, false positive LED candidates may appear, which are filtered out by the outlier rejection step.

The current Python-based implementation of the IRIS-filter and VBPC achieves a 2 Hz update rate. With the final C++ version, we expect real-time capabilities with both camera systems.

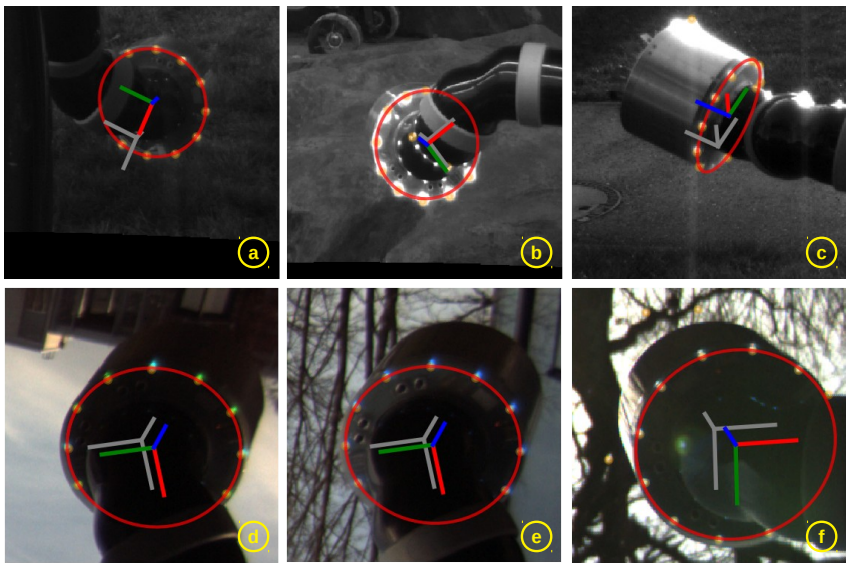


**Fig. 3** The end effector under varying illumination conditions with detected LED candidates (orange), the fitted ellipse (red), the center frame from the kinematics (gray), and the center frame by VBPC (RGB) from the navigation camera (top row), and from the rear camera (bottom row).

## 3 Accuracy

First, we evaluate the accuracy of our VBPC method. For this, we perform experiments with the LRU inside a laboratory environment. We use precise end effector

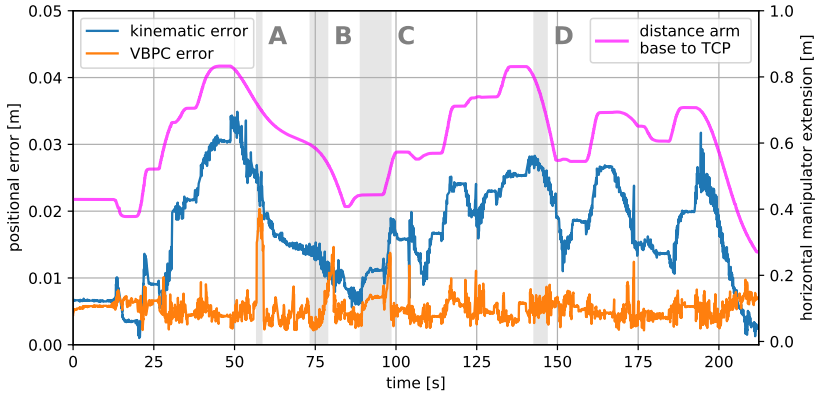


**Fig. 4** Results from the principal navigation camera: The kinematic position error and the VBPC position error w.r.t. the ground truth tracking system are shown. The kinematic error correlates with the horizontal extension of the manipulator (shown in magenta). The grey areas show periods of unsuccessful arm pose estimation by the VBPC.

tracking with a Vicon motion capture system to obtain a ground truth of the manipulator pose. The cameras are calibrated intrinsically and extrinsically and the images have been rectified.

During the experiments, the end effector is commanded to move to 15 different poses while being observed by the rover cameras. The target poses are chosen to cover the majority of the field of view of both cameras and to cover most of the manipulator's workspace.

The obtained positional accuracy of VBPC with the principal navigation camera is shown in Fig. 4 together with the accuracy of the forward kinematics, considering continuous movement of the manipulator. First of all, it can be seen that the positional error of the forward kinematics w.r.t. the ground truth strongly correlates with the horizontal extension of the manipulator, which is additionally shown as reference. The error of the forward kinematics can reach up to 3.5 cm at stretched out manipulator poses, which emphasizes the necessity for a vision based pose correction.

In contrast to the forward kinematics, the VBPC is able to determine the position of the end effector with a significantly higher accuracy, generally around 0.7 cm, infrequently reaching 1 cm. During the experiment, all false positives are rejected by our outlier detection step.

The VBPC fails to obtain valid measurements of the end effector position at four instances, which are marked grey in Fig. 4. The first two periods without valid updates (labeled A and B) are the result of an ill-posed ellipse fitting problem. During these periods, the marker-ring is observed by the navigation camera from the side, resulting in highly eccentric ellipses that also restrict the observability of the individual markers. In the two cases A and B, the eccentricity factors $\epsilon$ of the ellipses reach 0.99 or higher. The eccentricity factor is calculated as

$$\epsilon = 1 - \frac{a_{\text{minor}}}{a_{\text{major}}}, \tag{3}$$

where $a_{\text{minor}}$ and $a_{\text{major}}$ denote the length of the minor and major axis of the ellipse.

The other two periods without valid updates (marked as C and D), are due to partial or complete self-occlusions by the robotic arm (five or less markers are visible). Generally, as long as six or more LEDs are visible during the experiment, the VBPC is able to obtain a correct position measurement.

Regarding the orientation, VBPC is not able to correct the angular error. The average angular error of the forward kinematics lies at 3°, which can be considered as sufficiently accurate for manipulation. Recall that we use the orientation from the forward kinematics as initial guess for the iterative PnP approach. The angular error of the forward kinematics is too small to be further improved from the detected marker positions. Therefore, PnP is not able to improve the pose estimate compared to the initial guess and returns the corresponding initial values.

Finally, we evaluate the accuracy of VBPC considering the 15 static target poses observed by both the rear upwards facing camera and the principal navigation camera. For the principal navigation camera, the positional error is determined as $0.7 \pm 0.4$ cm for the 15 target poses. For the rear camera, the positional error is calculated as

0.5 cm with a standard deviation of 0.2 cm. The difference in accuracy between both cameras can be attributed to the different distances from the cameras to the end effector. Recall, that we evaluate the rear camera only at static poses, due to its low update rate and due to the manipulator frequently leaving the camera image during movement.

## 4 Experimental Robustness

To verify the robustness of the detection, we consider multiple experiments in indoor and outdoor scenarios. The goal is to cover the range of possible operation scenarios that the rover will face.

The first experiment (Exp-1) is the indoor scenario from Section 3 as reference. In general, indoor robot operation requires higher camera exposure time, causing an over-exposure of the LEDs, as for example illustrated in Fig. 3 (b). Therefore, Exp-1 can be used to evaluate the robustness w.r.t. over-exposed markers in the image.

With the outdoor scenarios, other aspects challenge the algorithm, which we reflect in the experiments accordingly. We consider different sun positions and the orientation of the rover with respect to the sun, as these cause different types and magnitudes of reflections and influence the appearance of the background. During the experiments, the robotic arm of the LRU moved to several poses, similar to the indoor experiment Exp-1.

The first outdoor experiment (Exp-2) considers a high standing sun. The remaining experiments were performed at low point sun positions (Exp-3 to 5). Examples of the marker appearance in the navigation cameras for high and low sun positions are shown in Fig. 3 (a) and (c) respectively and for the rear camera in (d) and (f).

For Exp-3, the rover is oriented with 90° to the sun. During Exp-4, the cameras face away from the sun, but the end effector appears fully illuminated in the camera image. Finally, the rover cameras face directly into the sun during the Exp-5. All selected experiments are listed in Table 1.

During all experiments, the manipulator is commanded to several target poses, where it remains static for 2 - 3 seconds and subsequently moves to the next pose. We evaluate the robustness of our detection regarding two criteria.

The first criterion is the capability to correctly detect the markers while the manipulator remains static at the respective target positions. This is the application that is most relevant for our future operations. Table 1 shows that the detection is successful in all five experiments, with only two exceptions during Exp-1 and Exp-3. In the case of the two unsuccessful detections, only five markers were visible, a correlation that was already described in Section 3. Indeed, the unsuccessful detection for Exp-1 is the grey area C in Fig. 4.

The other robustness evaluation criterion considers successful detection of the markers during continuous arm movement along varying trajectories. We consider every frame as a new pose, if the end effector moved more than 1 cm compared to the previously considered pose. The corresponding metric is the rate of successfully

**Table 1** Robustness experiments for the principal navigation camera. Experiments, where not all false positives are rejected by the outlier detection are marked with *.

| | Experiment type | end-effector movement [m] | static pose success rate | general rate of detection [%] | rate for $\epsilon < 0.9$ [%] |
|---|---|---|---|---|---|
| Exp 1 | indoor reference experiment | 4.0 | 14/15 | 86.5 | 92.1 |
| Exp 2 | outdoor, noon sun | 4.7 | 10/10 | 73.0 | 88.8 |
| Exp 3 | outdoor, low sun from 90° side | 4.65 | 9/10 | 90.2 | 92.3 |
| Exp 4 | outdoor, low sun from 180° | 1.4 | 5/5 | 47.2* | 60.8* |
| Exp 5 | outdoor, facing towards low sun | 4.0 | 10/10 | 76.4 | 88.4 |

detected poses compared to the overall number of poses. In Section 3, we established a connection that VBPC is prone to failure for highly eccentric ellipses, therefore we especially investigate the detection rate for well-conditioned ellipses ($\epsilon < 0.9$).

Table 1 shows that VBPC is able to detect the markers during approximately 90 % of each experiment considering ellipses with $\epsilon < 0.9$. A notable exception is Exp-4 with only 60 % success rate. During this experiment, the protective glass cover of the markers is frequently oriented perpendicular to the sun's direction, causing reflections that make the LEDs indistinguishable. This issue can be overcome by considering the sun orientation during robot operations or modifying the hardware using less reflective, matte materials.

Finally, we consider the upwards facing rear camera with the static poses. During Exp-1 to Exp-3, the end effector is reliably detected using VBPC in 14, 10, and 8 cases respectively. Contrary to this, during Exp-4 and Exp-5, VBPC fails to compute a correct pose estimate for most of the positions (it is only successful in 2/5 and 3/10 cases). This especially holds for end effector positions that appear in the upper parts of the image, where the bright background from the sky is combined with strong reflections. Furthermore, in these two cases, VBPC is susceptible to false positives that are not rejected by the outlier detection step, showing the limitations of our approach.

## 5 Conclusion and Outlook

We proposed a vision based pose correction to account for kinematic errors in robotic manipulators w.r.t. the observing cameras, which occur due to elasticities, extrinsic

camera decalibration, or other external influences. Our VBPC method considers a planetary-analog scenario and therefore focuses primarily on robustness towards outdoor illumination, and was specifically developed for the upcoming Moon-analog mission on the volcano Mt. Etna.

In our work, we showed the accuracy of the VBPC, which is able to significantly improve the pose estimate of the manipulator compared to the forwards kinematics. Our method is robust against external disturbances. This robustness especially holds for static pose corrections, but also allows to determine the manipulator pose during dynamic movements.

VBPC will be used as initialization for a tracking of the manipulator with the intended purpose to be used for a vision based servo control scheme. For future work, we additionally plan to use our detection for continuous camera calibration verification.

Furthermore, a tightly coupled fusion of forward kinematics and vision based observations is envisioned, where the pose estimation from vision and kinematics is coupled with the continuous parameter estimation of an error model.

## References

1. Breitenmoser, A., Kneip, L., Siegwart, R.: A monocular vision-based system for 6D relative robot localization. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 79–85 (2011)
2. Faessler, M., Mueggler, E., Schwabe, K., Scaramuzza, D.: A monocular pose estimation system based on infrared LEDs. In: IEEE International Conference on Robotics and Automation, pp. 907–913 (2014)
3. Fantacci, C., Pattacini, U., Tikhanoff, V., Natale, L.: Visual end-effector tracking using a 3D model-aided particle filter for humanoid robot platforms. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 1411–1418 (2017)
4. Garcia Cifuentes, C., Issac, J., Wüthrich, M., Schaal, S., Bohg, J.: Probabilistic articulated real-time tracking for robot manipulation. IEEE Robotics and Automation Letters **2**, 577–584 (2017)
5. Gratal, X., Romero, J., Kragic, D.: Virtual visual servoing for real-time robot pose estimation. In: IFAC World Congress, 1, pp. 9017 – 9022 (2011)
6. Kanatani, K., Sugaya, Y., Kanazawa, Y.: Guide to 3D Vision Computation. Springer (2016)
7. Klingensmith, M., Galluzzo, T., Dellin, C., Kazemi, M., Bagnell, J.A.D., Pollard, N.: Closed-loop servoing using real-time markerless arm tracking. In: IEEE International Conference on Robotics and Automation (Humanoids Workshop) (2013)
8. Kobatake, H., Hashimoto, S.: Convergence index filter for vector fields. IEEE Transactions on Image Processing **8**, 1029–1038 (1999)
9. Lambrecht, J.: Robust few-shot pose estimation of articulated robots using monocular cameras and deep-learning-based keypoint detection. In: International Conference on Robot Intelligence Technology and Applications, pp. 136–141 (2019)
10. Lehner, P., Brunner, S.G., Dömel, A., Gmeiner, H., Riedel, S., Vodermayer, B., Wedler, A.: Mobile manipulation for planetary exploration. In: IEEE Aerospace Conference, pp. 1–11 (2018)
11. Liang, C.J., Lundeen, K.M., McGee, W., Menassa, C.C., Lee, S., Kamat, V.R.: A vision-based marker-less pose estimation system for articulated construction robots. Automation in Construction **104**, 80–94 (2019)

12. OpenCV: Open source computer vision library (2020)
13. Ortenzi, V., Marturi, N., Stolkin, R., Kuo, J.A., Mistry, M.: Vision-guided state estimation and control of robotic manipulators which lack proprioceptive sensors. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 3567–3574 (2016)
14. Schmidt, T., Hertkorn, K., Newcombe, R., Marton, Z.C., Suppa, M., Fox, D.: Depth-based tracking with physical constraints for robot manipulation. In: IEEE International Conference on Robotics and Automation, pp. 119–126 (2015)
15. Schuster, M.J., Müller, M.G., Brunner, S.G., Lehner, H., Lehner, P., Dömel, A., Vayugundla, M., Steidle, F., Lutz, P., Sakagami, R., Meyer, L., Belder, R., Smisek, M., Stürzl, W., Triebel, R., Wedler, A.: Towards heterogeneous robotic teams for collaborative scientific sampling in lunar and planetary environments. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (Workshop on Informed Scientific Sampling in Large-scale Outdoor Environments) (2019)