

Robust Visual Localization for Hopping Rovers on Small Bodies

S. Chiodini¹, R.G. Reid², B. Hockman³, I.A.D. Nesnas², S. Debei¹ and M. Pavone³

Abstract—We present a collaborative visual localization method for rovers designed to hop and tumble across the surface of small Solar System bodies, such as comets and asteroids. In a two-phase approach, an orbiting primary spacecraft first maps the surface of a body by capturing images from various poses and illumination angles; these images are processed to create a prior map of 3D landmarks. In the second phase, a hopping rover is deployed to the surface where it uses a camera to relocalize to the prior map and to perform on-board visual simultaneous localization and mapping (SLAM). Small bodies present several challenges to existing visual SLAM algorithms, such as, high-contrast shadows that move quickly over the surface due to the short (e.g. 1-12 hour) rotational periods, and large changes in visual appearance between orbit and the surface, where image scale varies by many orders of magnitude (kilometers to centimeters). In this work, we describe how to augment ORB-SLAM2—a state of the art visual SLAM implementation—to handle large variations in illumination by fusing prior images with varying illumination angles. We demonstrate how a wide field of view (FOV) camera on a hopping rover can relocalize to prior maps captured by a narrow FOV camera taken from an orbiting spacecraft, and how the growth of pose and scale errors can be corrected through periodic loop closures during large hops. We evaluate the proposed method with sequences of images captured around a (1 meter) mock asteroid. The approach is shown to be robust to varying illumination angles, scene scale changes, and off-nadir camera pointing angles.

I. INTRODUCTION

The in-situ exploration of small Solar System bodies, such as asteroids and comets, is an emerging frontier of scientific interest, fundamental to improving our understanding of the formation of the early Solar System [1]. Conventional wheeled rovers are ineffective in microgravity environments due to lack of traction, and so new mobility paradigms, such as hopping and tumbling, are being investigated. To date, no robotic mobility systems have been successfully deployed to the surface of a small body; however, DLR’s MASCOTT rover and JAXA’s MINERVA 2 rovers are currently en route to Asteroid 162173 Ryugu aboard the Hayabusa 2 spacecraft [2].

The authors have been developing “Hedgehog”, a new internally-actuated hopping/tumbling rover concept.

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, and was sponsored by JPL Visiting Student Research Program and the National Aeronautics and Space Administration.

¹Sebastiano Chiodini and Stefano Debei are with CISAS, Università degli Studi di Padova, Padova, Italy {sebastiano.chiodini, stefano.debei}@unipd.it

²Robert G. Reid and Issa A.D. Nesnas are with Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA {rgreid, nesnas}@jpl.nasa.gov

³Ben Hockman and Marco Pavone are with the Department of Aeronautics and Astronautics, Stanford University, Stanford, CA, USA {bhockman, pavone}@stanford.edu

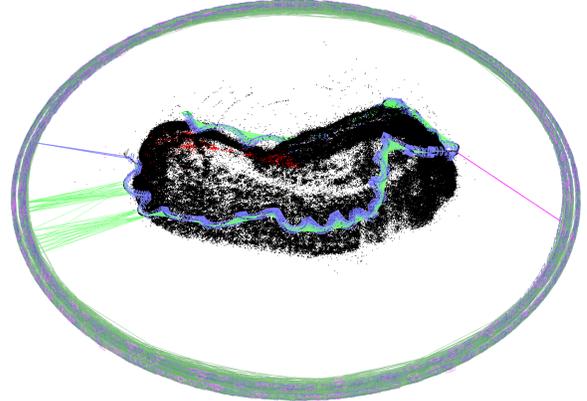


Fig. 1. Pose graph and 3D landmarks for a simulated mission to an asteroid. Blue frustums represent keyframes (camera images), green lines indicate the essential graph of landmark co-observations that connect keyframes, magenta lines indicate loop closure, black points are 3D landmarks on the surface of the body and red points are 3D landmarks that belong to the local map for the tracking. Here, the outer ring of highly-connected keyframes was obtained by a primary spacecraft in a prior mapping phase, while the hopping rover’s trajectory can be seen near the surface. The rover relocalizes itself to the prior map near the apogee of large hops, and these loop closures help to correct for accumulated errors via pose graph optimization.

Equipped with three orthogonal flywheels, Hedgehog is able to both controllably *tumble* by applying precise internal torques and also hop large distances by abruptly releasing stored energy in the flywheels via mechanical brakes. Recent experiments in various micro-gravity test beds have demonstrated the viability of Hedgehog’s mobility and control [3], [4], and a notional mission architecture is proposed in [5] and illustrated in Fig. 1.

On-board localization, which is required for autonomous surface activities, is particularly challenging for hopping rovers. On small bodies, vision-based localization must contend with (1) high-contrast, rapidly moving shadows that are characteristic of fast-rotating, airless bodies, (2) the large image scale variations associated with large, ballistic hops (centimeters to kilometers), (3) the inability to use accelerometers and stereo cameras to observe absolute scale, (4) detecting loop closures, or “relocalizing,” between wide-FOV images taken onboard the rover and nadir-pointing narrow-FOV images taken from the primary spacecraft, (5) lack of variety in surface features (e.g. rocks are often locally self-similar), (6) periods of rapidly rotating camera views as the rover tumbles in flight, and (7) frequent visual occlusions.

In this paper, we propose a collaborative visual localization method for hopping rovers that overcomes these challenges by adapting ORB-SLAM2, a state-of-the-art visual SLAM algorithm [6]. In a two-phase approach, a global map of

3D landmarks is first generated by an orbiting primary spacecraft with a narrow FOV camera, and this prior map is downloaded to the rover. After deployment to the surface, the rover uses a wide-FOV camera to perform on-board visual SLAM with reference to the prior global map. Specifically, the adaptations to ORB-SLAM2 presented in this paper demonstrate:

- Collaborative visual SLAM between a primary spacecraft and hopping rover in an asteroid-like environment,
- Wide-FOV cameras reliably relocalizing to a prior map generated by a narrow-FOV camera,
- Robustness to varying surface illumination angles,
- Robustness to large image scale changes, and
- Robustness to off-nadir camera pointing.

In Section II, we review related approaches for localization on small bodies and the state of the art in monocular visual SLAM. Section III describes how we adapted ORB-SLAM2 to handle the visual challenges faced by rovers hopping on small bodies. In Section IV, we evaluate the robustness and accuracy of our method with a series of experiments using realistic images of a mock asteroid with Sun-analog illumination and both narrow and wide FOV cameras executing orbital and hopping motions.

II. RELATED WORK

In recent years, several techniques have been proposed that combine visual and inertial sensors for the relative localization of hopping/tumbling rovers at the surface of small bodies. While the MINERVA rover never validated its localization approach, in [7], Yoshimitsu et al. describes how the rover was designed to estimate its attitude by fusing observations of the Sun and integrating gyroscope measurements. Relative velocities during hops were to be estimated using optical flow from surface images. In [8], Fiorini et al. propose a localization algorithm for hopping robots that estimates *a posteriori* trajectories and the landing area of a robot by fusing camera images and accelerometer, gyroscope, and contact sensor measurements that were recorded during jumps. Optical flow and visual odometry were explored by So et al. in [9] for the *relative* localization of a hopping rover. Their proposed algorithm works with tumbling camera motion and without continuous tracking of surface features. Estimates for the rover’s ballistic motion are computed from visual odometry captured at the start of each hop.

The unobservability of absolute metric scale using on-board sensors is a particularly insidious challenge for hopping rovers. Grounded rovers have two key advantages for scale estimation: the ability to fuse accelerometer data with visual odometry [10], and a continuous close-proximity vantage point to the surface for using stereo cameras—both of which are largely unavailable for hopping rovers during large ballistic (“acceleration-free”) trajectories. In [9], So et al. describe how a stereo pair of cameras can help recover absolute scale at the start of each hop, but the small baselines afforded by hopping rovers (roughly 10 cm) provide very little information when hops exceed

10 meters in height. Thus, for the vision-based techniques described in the literature, hopping rovers are highly prone to accumulating errors in position and scale, or “drift.”

However, “loop closures,” or the ability to recognize the same place when revisited, allows visual SLAM algorithms to *correct* for drift. A thorough review of the state of the art in visual SLAM can be found in [11]. In general, early approaches estimated pose and landmarks online using extended Kalman filters, while newer approaches use incremental batch optimization techniques such as bundle adjustment [12] and pose graph optimization [13].

ORB-SLAM2 [6], the visual SLAM algorithm modified for this work, integrates both bundle adjustment and pose graph optimization approaches to create an implementation that is robust and efficient. ORB-SLAM2 uses Lie group $\text{Sim}(3)$ to represent *constraints* in its pose graph optimization, which allows it to explicitly parameterize and correct for scale drift. For robust tracking and estimation, ORB-SLAM2 uses Oriented Binary (ORB) features [6] to detect and track landmarks in camera images; these ORB features require minimal storage and are efficient to calculate, making them well-suited for space applications, which often have limited computation. ORB features describe landmarks with 3D position estimates that are compressed into *words* (here integers) and placed in a *bag-of-words* database that is used to quickly detect loop closure hypotheses [14].

III. PROPOSED METHOD

This section describes how to adapt the ORB-SLAM2 algorithm to enable collaborative localization for a typical mission scenario to a small body. Refer to Fig. 2a for an overview of mission phases. During a prior mapping phase, the primary spacecraft images the body as it rotates in view. Over many weeks or months, the spacecraft maneuvers to image the entire surface from all longitudes, θ_a , and from multiple relative surface illumination angles, α_s . After deployment to the surface, the rover hops and tumbles, accumulating position and scale errors (the rover’s attitude is constrained by gyroscope sensors and occasional observations from a star tracker).

To reduce these errors, the rover occasionally performs large hops to capture views of the surface that can be matched to the prior map (see Fig. 2a). Relocalizing to the prior map creates loop closure constraints in the pose graph that, after maximum likelihood estimation, correct the rover’s on-board pose estimates and reduce its pose uncertainty.

Both the primary spacecraft’s and rover’s maps are composed of sets of keyframes K_j and landmarks X_l . Each keyframe stores the $\text{SE}(3)$ pose relative to the world frame ${}^w\mathbf{T}_j$, the camera intrinsic, the ORB descriptors and their image positions $\mathbf{x}_{j,l}$, which are possibly associated with landmarks. Each landmark includes its world position ${}^w\mathbf{X}_l$, the average viewing direction from all observations, the ORB descriptor, and the maximum d_{max} and minimum d_{min} distances at which the point is likely to be observed. Landmarks can also be augmented with an estimate of the

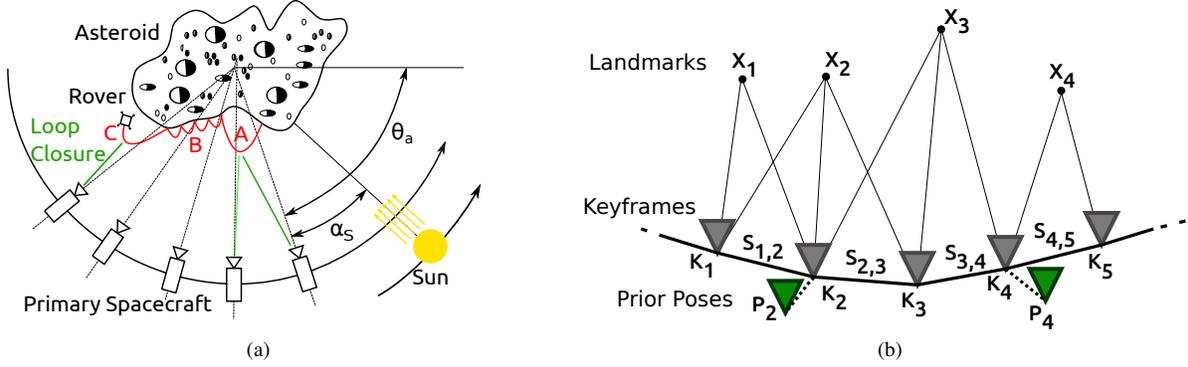


Fig. 2. (a) In a prior mapping phase, the primary spacecraft images the body as it rotates in view. Over many weeks or months, the spacecraft maneuvers slowly to image the entire surface. Throughout this paper, longitude around the body’s rotational axis is denoted θ_a , while the relative surface illumination angle is denoted α_s . After the rover is deployed to the surface (red) its local illumination angle continues to vary from sunrise ($\alpha_s \simeq -90^\circ$) to noon ($\alpha_s = 0^\circ$) and to sunset ($\alpha_s \simeq 90^\circ$). (b) Graphical model for the prior mapping phase. K_j are keyframes, P_k are keyframe priors (e.g. ground-based pose estimates), X_l are 3D landmarks and $S_{j,j+1}$ are measurements of the relative Sim(3) transformation between keyframes.

local illumination angle α_s from when they were observed. An example of estimating landmark pose uncertainties in a real mission is given in [15].

A. Prior mapping by the primary spacecraft

In a typical mission to a small body, a navigation team on the ground uses additional measurements such as radio doppler to estimate the primary spacecraft’s orbital trajectory. These measurements facilitate the estimation of absolute scale and, when combined with a star tracker, can also provide SE(3) priors on the spacecraft’s pose during the creation of the global map. These priors are added every n camera frames, as shown in the graphical model in Fig. 2b. The estimation of a spacecraft’s trajectory, or “orbit determination,” can achieve astonishing precision; for example, the 1σ uncertainties in attitude and position of Rosetta spacecraft with respect to the comet at Philae separation were 0.015° and 10 m [16].

On the primary spacecraft, a *prior adjustment* thread has been added to ORB-SLAM2 that promotes camera images with SE(3) priors to keyframes and performs additional pose graph optimizations. These pose graph optimizations are performed over the *essential graph*. The essential graph connects keyframes’ poses with estimated transformations, or graph constraints. These constraints are either derived by minimizing the reprojection error of the map points with respect to the matched keypoints (e.g. in the tracking front-end or during relocalization) or by finding a similarity transformation using the method of Horn (e.g. in the loop closing). These constraints are effectively derived from large sets ORB features that have been co-observed between keyframes, and they are represented by a Sim(3) Lie group. $\mathfrak{sim}(3)$ is the corresponding Lie algebra, represented by a 7-vector $(\boldsymbol{\omega}, \mathbf{v}, \sigma)$, where $\boldsymbol{\omega} = (\omega_1, \omega_2, \omega_3)$ is the axis-angle representation of the rotation, \mathbf{v} is the rotated version of the translation \mathbf{t} and $\sigma = \log s$, s is the scale [17]. The relation

between $\mathfrak{sim}(3)$ to Sim(3) is given by

$$\begin{pmatrix} \boldsymbol{\omega} \\ \mathbf{v} \\ \sigma \end{pmatrix} = \log \text{Sim}(3) \begin{bmatrix} s\mathbf{R} & \mathbf{t} \\ 0 & 1 \end{bmatrix}. \quad (1)$$

The pose graph optimization distributes residual errors between the various constraints in the essential graph. A residual is defined as

$$\mathbf{e}_{j,j+1} = \log \text{Sim}(3)({}^w\mathbf{S}_{j,j+1} {}^w\mathbf{S}_{j+1,w} {}^w\mathbf{S}_{j,w}^{-1}), \quad (2)$$

where ${}^w\mathbf{S}_{j,j+1}$ is the relative Sim(3) transformation between the connected keyframes computed before the pose graph optimization expressed in the world reference frame w , and ${}^w\mathbf{S}_{j,w}$ is the Sim(3) transformation between the frame j and the world reference frame.

During the pose graph optimization, a “virtual” constraint is added between the tracked pose K_j and the prior pose P_k , and the relative transformation between them is set to identity, ${}^w\mathbf{S}_{j,k} = \mathbf{I}$. The constraint error between a tracked keyframe and its prior pose is defined as

$$\mathbf{e}_{j,k} = \log \text{Sim}(3)(\mathbf{I} {}^w\mathbf{P}_{k,w} {}^w\mathbf{S}_{j,w}^{-1}), \quad (3)$$

where ${}^w\mathbf{P}_{k,w}$ is the transformation between the prior pose and the world reference frame. The cost function in the ORB-SLAM2 essential graph optimization is augmented with the pose priors,

$$\chi^2({}^w\mathbf{S}_{2,w}, \dots, {}^w\mathbf{S}_{m,w}) = \sum_j \mathbf{e}_{j,j+1}^\top \mathbf{e}_{j,j+1} + \sum_j \mathbf{e}_{j,k}^\top \Omega_{\Delta S_{j,k}} \mathbf{e}_{j,k}, \quad (4)$$

where $\Omega_{\Delta S_{j,k}}$ is the diagonal information matrix for each prior, $\Omega_{\Delta S_{j,k}} = \text{diag}(\sigma_r^2, \sigma_r^2, \sigma_r^2, \sigma_t^2, \sigma_t^2, \sigma_t^2, \sigma_s^2)$. Here, σ_r , σ_t , σ_s are the priors’ rotation, translation and scale standard deviation estimates, respectively. These values are manually tuned to ensure pose graph convergence as they are sensitive to the scene scale and camera parameters (FOV, resolution etc.). After the essential graph optimization, keyframe poses are adjusted, and a further global bundle adjustment is performed to refine the keyframe poses and landmark estimates.

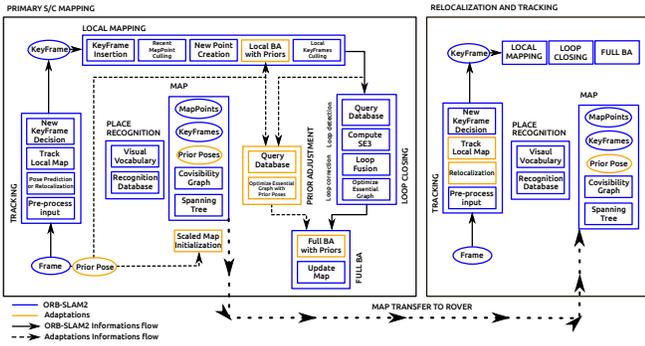


Fig. 3. Block diagram of the adaptations made to ORB-SLAM2 to enable robust visual localization for tumbling rovers. Based on ORB-SLAM2 diagram [6].

B. Rover relocalization on the prior map

The hopping rover’s relocalization approach is very similar to that of ORB-SLAM2; first, a *bag-of-words* technique provides a set of candidate keyframe matches via a fast database search [14]. These candidate keyframes are filtered using two parameters, k_{words} and k_{score} , which provide a lower bound on the number of shared words (ORB features), and a similarity score, respectively. For each keyframe candidate that passes the filters, the ORB image features are matched to the current frame. A geometric consistency test is then performed between the 2D image features and the 3D landmarks using RANSAC and a PnP algorithm. If enough inliers are found, the current frame is considered “relocalized” and the relative pose is further optimized by considering all inliers. A search is performed via the essential graph to establish additional feature/landmark correspondences, and the keyframes and landmarks are optimized using bundle adjustment.

The following measures have been adopted to make relocalization more robust (parameters have been tuned according to our experimental set-up). (1) By increasing the number of pyramid levels of the ORB feature extractor during mapping, it is possible to relocalize the rover from the beginning of its descent. Since the spatial resolution of the wide FOV images is far larger (6.2 times greater) than that of the narrow FOV images, 11 pyramid levels are needed in the narrow FOV feature extraction to match the same feature scale as the wide FOV features (the scale difference between ORB adjacent pyramid levels is 1.2). (2) By reducing the minimum number of inliers in the RANSAC test, it is possible to increase the number of relocalized frames near the asteroid surface. As the relocalization occurs in correspondence to the features presenting the finest scale, only a sub-set of the extracted features is usable. Assuming 4000 features are extracted, only 385 belong to the first pyramid level, and if the rover is in a 100 m hop while the primary spacecraft is orbiting at 1000 m altitude only 1/16.5 of the first pyramid level features are visible from the rover’s wide FOV camera. For these reasons, the threshold on the minimum number of inliers is decreased from 50 to 10. (3) By disabling scale filtering, it is possible to track new wide FOV frames after relocalization.

IV. EXPERIMENTAL EVALUATION

To demonstrate the proposed adaptations to ORB-SLAM2, a series of experiments were performed around a 1.4 meter diameter mock-asteroid in the JPL Robodome. To simulate realistic illumination changes, a collimated light was placed 5 m away from the asteroid and the asteroid was mounted so it could rotate around a single axis. Three configurations were tested: (1) both prior mapping (primary spacecraft) and visual SLAM (deployed rover) with static illumination, (2) prior mapping with the asteroid rotating (changing illumination) and visual SLAM with fixed illumination and (3) both prior mapping and visual SLAM with the asteroid rotating (most realistic). The prior mapping phase was performed with the primary spacecraft orbiting close to the asteroid’s equatorial plane. While the mock asteroid had a diameter of 1.4 m, in order to compare the results with real exploration missions, all measurements have been normalized here to an asteroid with an effective diameter of 1000 m, which is around the estimated size of 162173 Ryugu asteroid. A video attachment of the experiments is available at <https://youtu.be/acI0axPgJCY>.

The rover’s wide FOV camera had a FOV of $100^\circ \times 82^\circ$ and a resolution of 640×480 pixels. The primary spacecraft’s narrow FOV camera had a FOV of $60^\circ \times 35^\circ$ and 1920×1080 pixels resolution. Both cameras were calibrated with Zhang’s calibration method [18]. The motion of both cameras and the asteroid were tracked using a Vicon motion capture system, which provided a 2 mm tracking accuracy.

In order to analyze the algorithm’s performance and generate the prior map, the timestamps of the camera and the Vicon system must first be synchronized. Moreover, the two trajectories are captured in two different reference frames: the Vicon reference frame w and the camera reference frame c . The ORB-SLAM2 trajectory is expressed with reference to the first keyframe pose. In order to align the two timestamps, we first estimate the time-shift by identifying some distinct feature in the time-displacement plot. Then, the Vicon data is interpolated at the camera timestamp values. With the two sets of trajectory data now synchronized, we then estimate the rotation matrix ${}^w_c \mathbf{R}$, translation ${}^w \mathbf{t}_{c,w}$, and scale $s_{c,w}$ between the camera and the Vicon reference frames, and translation between the camera center and the camera rig, ${}^r \mathbf{t}_{c,r}$, by solving the following non-linear optimization problem:

$$\underset{({}^w_c \mathbf{R}, {}^w \mathbf{t}_{c,w}, s_{c,w}, {}^r \mathbf{t}_{c,r})}{\operatorname{argmin}} \sum_i \| {}^w \mathbf{r}_i - s_{c,w} {}^w_c \mathbf{R} {}^c \mathbf{r}_i - {}^w \mathbf{t}_{c,w} + {}^w \mathbf{R} {}^r \mathbf{t}_{c,r} \|^2, \quad (5)$$

where ${}^w \mathbf{R}$ is the pose of the camera rig in the Vicon frame of reference.

A. Performance in an asteroid-like environment

The images required for the prior mapping were captured by manually moving the narrow FOV camera, mounted on a cart, around the asteroid in a circle (i.e. an “orbit”). The light direction was kept fixed parallel to the asteroid orbital plane. Vicon data was used to create priors. Two mapping sequences were captured: the *far sequence map*, which was taken at a

distance of 160 cm from the asteroid (see Fig. 4), and the *three distance map*, which was the result of three consecutive sequences closer and closer to the asteroid (160 cm, 120 cm and 80 cm, respectively).

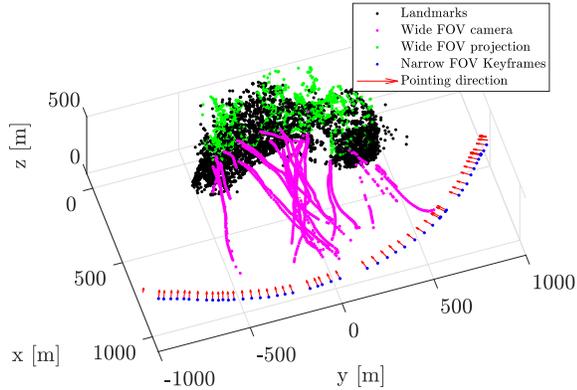


Fig. 4. Successful rover relocalizations in the *far sequence map*. Keyframes from the primary spacecraft’s prior map are shown in blue, with red vectors indicating the camera’s optical axis. Dozens of simulated descents to the surface were performed in this image sequence; magenta markers indicate where single images from the rover’s wide FOV camera were successfully relocalized in the primary spacecraft’s prior narrow-FOV images. The projection of the camera’s optical axis onto the body (green) allows the camera’s off-nadir pointing angle to be estimated. Distances are normalized to a 1000 m asteroid.

Fig. 5 shows the wide FOV relocalization rates over prior maps which were generated by setting the following tuning parameters: prior insertion rate n , information matrix values σ_r , σ_t and σ_s , and ORB features extracted for each image N . The absolute trajectory estimation error for the narrow FOV mapping is 3 m, while the accuracy of the (normalized) motion capture ground truth is 1.4 m. These values have the same order of magnitude as the accuracy with which we can estimate the pose of the spacecraft using orbit determination [19]. The angular error is only 0.2° along each axis.

B. Relocalizing a wide FOV camera to narrow FOV priors

The robustness of the algorithm to relocalize the wide FOV camera to narrow FOV priors was tested with a sequence of images simulating dozens of wide FOV camera descents to the asteroid mock up surface (again, with a fixed light direction). The wide FOV camera was moved by hand to simulate the rover’s descent to and hops on the asteroid surface. The reference trajectory was extracted by tracking four retro-reflective markers on the camera. Despite the large scale variations encountered during the drastic “in-and-out” motions, Fig. 4 shows that about 40% of the 6100 wide FOV images were relocalized. An example of a successfully relocalized frame is shown in Fig. 6.

Relocalization rates were evaluated by varying the number of map candidate keyframes returned by the BoW library, changing the two parameters k_{words} and k_{score} , and by modifying the number of iterations performed in the RANSAC scheme during the geometric consistency test. We focus on the wide FOV relocalization rate for a rover hop height of 150 m. As we can see from Fig. 5, by filtering the candidate

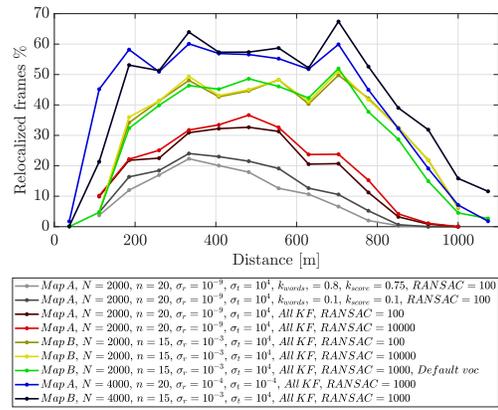


Fig. 5. Percentage of successful rover relocalizations vs. the normalized distance from the asteroid surface for the simulated descents shown in Fig. 4. These histograms indicate the percentage of the rover’s wide-FOV camera images that were successfully relocalized in the primary spacecraft’s narrow FOV image priors in the *three distance map* (Map A) and *far sequence map* (Map B). For both map priors, the best relocalization rates occurs when the number of ORB features per frame is increased to $N = 4000$. The parameters used in the bag-of-words keyframe candidate selection and RANSAC geometry consistency tests are also varied. Distances are normalized to a 1000 m asteroid.

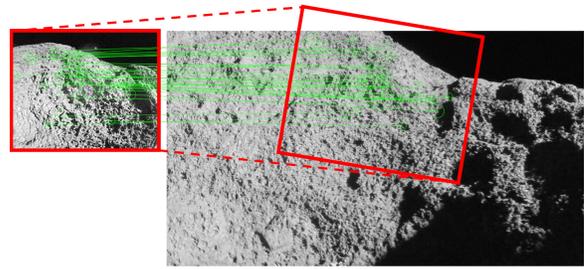


Fig. 6. Example relocalization. Left: an image from a hopping rover’s low-resolution wide-FOV camera (640×480 pixels, $f = 274$ pixels). On the right, this image has been successfully matched to a high-resolution narrow-FOV camera image (1920×1080 pixels, $f = 1527$ pixels) from the primary spacecraft’s prior map.

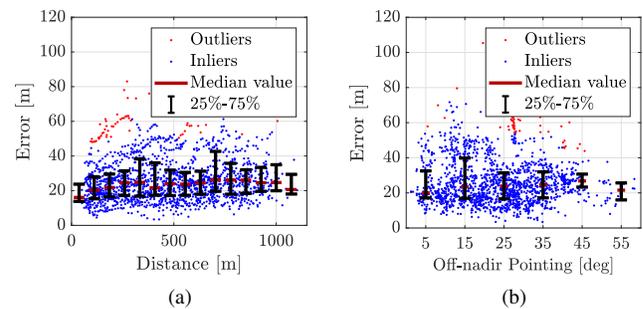


Fig. 7. Relocalization error for the rover’s wide-FOV camera in the simulated descents shown in Fig. 4 and the *three distance map* priors with $n = 20$ and $N = 4000$. Distances are normalized to a 1000 m asteroid. (a) Relocalization error vs. the normalized distance to the asteroid surface. (b) Relocalization error vs. the off-nadir pointing directions, which occurs when the rover tumbles.

map keyframes, the relocalization rate is 10%, which grows up to 17% if we take all the keyframes in the map. The

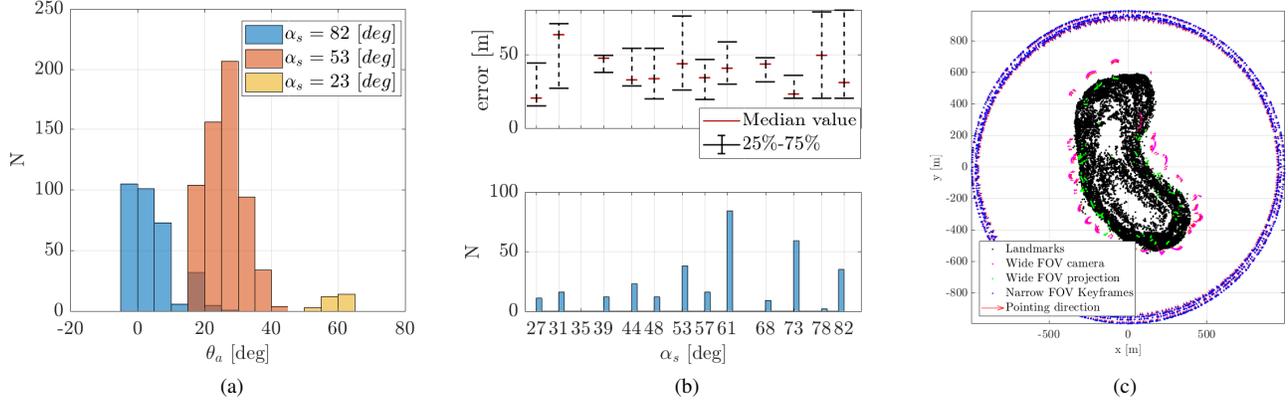


Fig. 8. (a) Histogram of successful wide-FOV camera relocalizations in the simulated descents (from Fig. 4) vs. the rover’s longitudinal position θ_a above the asteroid’s surface for three prior maps. The three prior maps were captured with different surface illumination angles as measured between the primary spacecraft and the sun while the asteroid rotated ($\alpha_s = 23^\circ$, 53° and 82°). The wide-FOV simulated descents were recorded with a fixed $\alpha_s = 82^\circ$. (b) Wide FOV camera relocalization on a large map generated by fusing individual maps illuminated every $\alpha_s = 5^\circ$. Top: relocalization errors as function of α_s . Bottom: the number of localized frames for each α_s map. (c) The final map and successfully relocalized wide FOV frames.

relocalization rate at 150 m is 38% if the *far distance map* is built by extracting twice as many (4000) ORB features for each image. By using the same map parameters to build the *three distance map*, the rate grows up to 52%.

Fig. 7a illustrates the accuracy of the localization algorithm as a function of the distance from the asteroid. The mean localization accuracy ranges from 16 m to 26 m.

C. Robustness to off-nadir camera pointing

The localization accuracy of the wide FOV camera on the prior map was evaluated as a function of the off-nadir pointing direction, which is the angle difference between the optical axis of the localized frame and the optical axis of the corresponding keyframe in the reference map. As we can see from Fig. 7b, ORB-SLAM2 can handle misalignments up to about 45 degrees.

D. Robustness to illumination changes

To test the algorithm’s robustness to variations in surface illumination angle, a series of maps was created by changing the direction of the light source, and the wide FOV camera was relocalized over them. Each of these maps was generated by mounting the narrow FOV camera on a tripod while the asteroid was rotated with respect to a fixed light source, whereby the light source was relocated for each map. This allowed the formation of perfectly circular mapping orbits around the asteroid.

Fig. 8a shows a histogram of successfully relocalized wide FOV frames as function of the longitude around the body’s rotational axis θ_a for different relative surface illumination angles α_s . The wide-FOV simulated descents were recorded with a fixed $\alpha_s = 82^\circ$, thus, when $\theta_a = 82^\circ - \alpha_s$ the wide FOV frame had the same lighting conditions as the map keyframe. The data in Fig. 8a suggests a robustness to lighting angle of about $\Delta\theta_a = 15$ degrees.

While ORB features are relatively robust to illumination angle ($\pm 15^\circ$), the ability to localize at *any* time of day

(i.e. $\pm 180^\circ$) still requires a prior map with images from a wide variety of illumination angles. To this end, a map was generated by fusing a series of maps with the lighting angle $\Delta\alpha_s$ varying in increments of 5 degrees. Fig. 8b shows the number of relocalized frames for each illumination condition and the corresponding average relocalization errors. The data suggests that some illumination angles are more favorable than others for mapping. For example, if α_s is close to 0° , the sun is behind the camera and very few shadows (and thus, features) are visible. On the other hand, if α_s is close to 90° , large portions of the asteroid surface are occluded in shadow, also limiting the number of features. The “sweet spot” for abundant features seems to be about 50° to 70° .

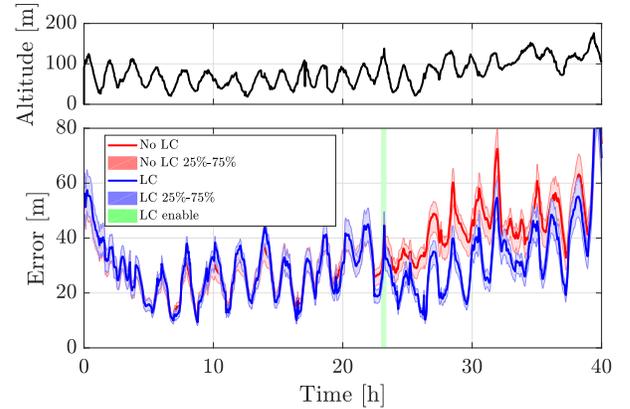


Fig. 9. Instantaneous pose errors for the long hopping sequence shown in Fig. 1, where the rover uses its wide FOV camera to relocalize multiple times to the prior map while hopping across the mock asteroid’s surface. Top: wide FOV camera distance from the the mock asteroid’s surface. Bottom: instantaneous pose errors, the red line represents the visual odometry results (no loop closure) and the blue line represents the visual SLAM results (loop closure enable at $t = [23.23.35]$ h interval). The randomness results from the RANSAC scheme are averaged over 10 runs. Starting from $t = 23.35$ h without loop closure the error drifts over time. Distances are normalized to a 1000 m asteroid.

Once the rover relocalizes over the prior map, the adapted algorithm is able to do visual SLAM on the asteroid surface and create loop closure constraints after large hops. Figure 9 demonstrates this capability, where the visual SLAM tracking error with loop closure *enabled* (at $t = [23\ 23.35]$ h interval) corrects for about 20 m of drift compared to the tracking error with loop closure *disabled*. At $t = 39$ h, even though most of the asteroid surface is covered in shadows, and the error grows, ORB-SLAM2 keep wide FOV camera pose track.

V. CONCLUSIONS

In this paper, we present a collaborative visual localization framework for a rover designed to hop and tumble across the surfaces of small solar system bodies. The proposed method is adapted from the state-of-the-art ORB-SLAM2 algorithm, and is based on a two-phase approach: first, the primary spacecraft maps the surface of a small body with its narrow FOV camera. Then, the rover is deployed to the small body surface and performs visual SLAM with its wide FOV camera. We test the method using data collected around a mock asteroid, where camera pose estimates are remarkably consistent with ground truth despite the occasional “feature-less” appearance of the asteroid.

We demonstrate the ability to relocalize a wide FOV camera over a prior map captured by a narrow FOV camera, and the ability to perform visual SLAM over the asteroid surface by creating loop closure constraints during large hops. When compared to ground truth motion capture data and normalized to a reference scale of 1000 meters, errors in trajectory estimates average 25 meters. It is important to note however, that this normalized error is primarily due to the focal length limitations of our mapping camera and the inability of our wide FOV camera to get “very close” to the surface. In practice (i.e. for a *true* 1000 m asteroid), the rover may be able to localize much more precisely.

Experimental results suggest that the algorithm is robust to large variations in scale, viewing angle, and lighting angle. Relocalization occurs when the distance to the asteroid surface is between 1087 m and 37 m, when the viewing angle misalignment is less than 45° , and when the difference in lighting direction is less than 15° . Finally, by fusing together prior maps generated from a range of lighting angles, we show that a rover can reliably localize at any time of day.

There are several ways in which this work can be extended in the future. For one, this study assumed that the hopping rover is able to use reaction wheels to stabilize its attitude during hops. Future work will address the case where the rover may be *tumbling*—a notable challenge since visual features continuously rotate into and out of view. Also, this study does not adequately assess the performance of visual localization at varying *latitudes*, and in particular, the challenges that may arise when attempting to relocalize near polar regions for a map generated solely from an equatorial vantage point. Finally, future studies should address the synergies that onboard visual localization may have with other mothership-in-the-loop methods such as radar Doppler and ranging.

ACKNOWLEDGMENTS

The authors wish to thank JPL’s Formation Control Testbed team (Carl Seubert, Kevin Lo and Dylan Conway) and Thomas Paynter, who assisted with the experiments.

REFERENCES

- [1] *Decadal survey vision and voyages for planetary science in the decade 2013-2022*. Technical report, National Research Council, 2011.
- [2] S. Ulamec, V. Kucherenko, J. Biele, A. Bogatchev, A. Makurin, and S. Matrossov, “Hopper concepts for small body landers,” *Advances in Space Research*, vol. 47, no. 3, pp. 428 – 439, 2011.
- [3] B. Hockman, A. Frick, R. G. Reid, I. A. Nesnas, and M. Pavone, “Design, control, and experimentation of internally-actuated rovers for the exploration of low-gravity planetary bodies,” *Journal of Field Robotics*, vol. 34, no. 1, pp. 5–24, 2017.
- [4] B. Hockman, R. G. Reid, I. A. D. Nesnas, and M. Pavone, “Experimental methods for mobility and surface operations of microgravity robots,” in *Int. Symp. of Experimental Robotics*, Oct. 2016.
- [5] M. Pavone, J. C. Castillo-Rogez, I. A. D. Nesnas, J. A. Hoffman, and N. J. Strange, “Spacecraft/rover hybrids for the exploration of small solar system bodies,” in *2013 IEEE Aerospace Conference*, March 2013, pp. 1–11.
- [6] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, “ORB-SLAM: A versatile and accurate monocular slam system,” *IEEE Transactions on Robotics*, vol. 31, no. 5, pp. 1147–1163, Oct 2015.
- [7] T. Yoshimitsu, T. Kubota, I. Nakatani, T. Adachi, and H. Saito, “Micro-hopping robot for asteroid exploration,” *Acta Astronautica*, vol. 52, no. 26, pp. 441 – 446, 2003.
- [8] P. Fiorini, C. Cosma, and M. Confente, “Localization and sensing for hopping robots,” *Autonomous Robots*, vol. 18, no. 2, p. 185, 2005.
- [9] E. W. Y. So, T. Yoshimitsu, and T. Kubota, “Hopping odometry: Motion estimation with selective vision,” in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Oct 2009.
- [10] A. I. Mourikis and S. I. Roumeliotis, “A multi-state constraint Kalman filter for vision-aided inertial navigation,” in *Robotics and Automation, 2007 IEEE International Conference on*, 2007, pp. 3565–3572.
- [11] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard, “Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age,” *IEEE Transactions on Robotics*, vol. 32, no. 6, pp. 1309–1332, Dec 2016.
- [12] G. Klein and D. Murray, “Parallel tracking and mapping for small ar workspaces,” in *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, Nov 2007, pp. 225–234.
- [13] K. Konolige, J. Bowman, J. D. Chen, P. Mihelich, M. Calonder, V. Lepetit, and P. Fua, “View-based maps,” *The International Journal of Robotics Research*, 2010.
- [14] D. Galvez-Lpez and J. D. Tardos, “Bags of binary words for fast place recognition in image sequences,” *IEEE Transactions on Robotics*, vol. 28, no. 5, pp. 1188–1197, Oct 2012.
- [15] S. Debei, A. Aboudan, G. Colombatti, and M. Pertile, “Lutetia surface reconstruction and uncertainty analysis,” *Planetary and Space Science*, vol. 71, no. 1, pp. 64 – 72, 2012.
- [16] J. Biele *et al.*, “The landing(s) of philae and inferences about comet surface mechanical properties,” *Science*, vol. 349, no. 6247, 2015.
- [17] H. Strasdat, J. Montiel, and A. J. Davison, “Scale drift-aware large scale monocular SLAM,” *Robotics: Science and Systems VI*, 2010.
- [18] Z. Zhang, “A flexible new technique for camera calibration,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 11, pp. 1330–1334, Nov 2000.
- [19] L. Lorda, E. Canalias, T. Martin, R. Garmier, and J. Biele, “Mascot: Analyses of the descent and bouncing trajectories to support the landing site selection,” in *26th International Space Symposium in Flight Dynamics, Matsuyama, Japan*, 2017.