Precision Instrument Targeting via Image Registration for the Mars 2020 Rover

Gary Doran, David R. Thompson, Tara Estlin
Jet Propulsion Laboratory, California Institute of Technology
{Gary.B.Doran.Jr,David.R.Thompson,Tara.A.Estlin}@jpl.nasa.gov

Abstract

A key component of Mars exploration is the operation of robotic instruments on the surface, such as those on board the Mars Exploration Rovers, the Mars Science Laboratory (MSL), and the planned Mars 2020 Rover. As the instruments carried by these rovers have become more advanced, the area targeted by some instruments becomes smaller, revealing more fine-grained details about the geology and chemistry of rocks on the surface. However, thermal fluctuations, rover settling or slipping, and inherent inaccuracies in pointing mechanisms all lead to pointing error that is on the order of the target size (several millimeters) or larger. We show that given a target located on a previously acquired image, the rover can align this with a new image to visually locate the target and refine the current pointing. Due to round-trip communication constraints, this visual targeting must be done efficiently on board the rover using relatively limited computing hardware. We employ existing ORB features for landmark-based image registration, describe and theoretically justify a novel approach to filtering false landmark matches, and employ a random forest classifier to automatically reject failed alignments. We demonstrate the efficacy of our approach using over 3,800 images acquired by Remote Micro-Imager on board the "Curiosity" rover.

1 Introduction

A key challenge in the operation of spacecraft on the surface of other bodies in the solar system is the low-bandwidth, high-latency communications imposed by the extreme distances between the spacecraft and Earth. Even at Mars, for which satellites are used as relays to increase the communication bandwidth between rovers and Earth, there is rarely the opportunity for more than one command uplink cycle per sol (Mars day). For future proposed missions to bodies in the outer solar system, such as the moons of Jupiter and Saturn, communications will be even more limited. Hence, to maximize the useful scientific data that is collected by these missions, there is need to develop spacecraft that are more capable of precise measurements without direct supervision.

A straightforward part of making spacecraft more capable is ensuring that they take measurements of the intended targets selected by humans. For example, scientists on the ground might identify an interesting rock in a rover's navigation camera image and command a follow-up observation with another instrument on board. The command might specify the intended target as a position in the local coordinate system. However, an accumulation of small errors in the estimated target position from stereo vision, the rover's position due to slipping or thermal variations, and the instrument's pointing due to imperfect mechanisms might lead a narrow-field-of-view instrument to miss the intended target. Clearly, a human on the surface of Mars would not make such a mistake; they would ensure that the target of interest from the navigation camera image was visible in the field-of-view of the instrument before taking a measurement.

The purpose of this paper is to describe a system for precision instrument targeting through "pointing refinement" that is aimed at mimicking the process of a human pointing an instrument at a target of interest. Instead of specifying a target as just a point in space, the target is also specified as a point within a context image. Then, a visual alignment of the context image and the instrument's field-of-view is used to verify that the instrument is pointed sufficiently close to the intended target. If not, the pointing is adjusted to bring the instrument into alignment with the intended target. The capability to refine pointing visually enables measurements for a whole new class of small-scale features that could not be targeted otherwise and are important for investigating biological signatures or assessing a target for sample return selection.

Below, we describe a landmark-based image registration approach to perform precision targeting. Although our method can be extended to other spacecraft, we specifically focus on pointing refinement for the future Mars 2020 rover mission, using data from the Mars Science Laboratory (MSL) "Curiosity" rover as an analog. In particular, we focus on pointing refinement for the "SuperCam" instrument selected for inclusion on Mars 2020 [Maurice *et al.*, 2015]. SuperCam is an upgraded version of the ChemCam instrument on board MSL [Maurice *et al.*, 2012; Wiens *et al.*, 2012]. First, we describe a novel strategy for operating instruments with precision targeting and compare this to current operations and prior work. Then, we describe innovations in density-based outlier removal and classification-based post-processing that

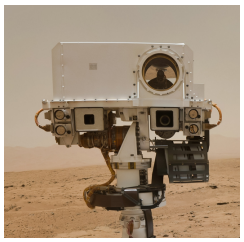

Figure 1: The ChemCam instrument is located on the mast of the Mars Science Laboratory "Curiosity" rover, where its large circular aperture is visible.

enable fast, robust matching of challenging natural rock surfaces using landmark-based image registration. Finally, we demonstrate the efficacy and feasibility of the pointing refinement using images from MSL's ChemCam instrument.

2 Overview

The ChemCam (short for "Chemistry and Camera") instrument on board the MSL rover (see Figure 1) uses a laser-induced breakdown spectrometer (LIBS) to remotely measure the composition of a rock target by firing a powerful laser at the rock [Maurice *et al.*, 2012; Wiens *et al.*, 2012]. The Remote Micro-Imager (RMI) within ChemCam is a telescopic camera with a narrow $19\,\mathrm{mrad}$ field-of-view, used to determine the context and position of the LIBS laser shot, which occurs near the center of the field-of-view. Figure 2 shows example RMI images, with red crosshairs indicating the position of the LIBS shot.

A simplified example of how ChemCam is operated on Mars is shown in the upper half of Figure 2 for a rock target called "McGrath." On Sol N, an RMI-only image of McGrath was acquired for the purposes of carefully pointing the LIBS laser at a vein in the image, indicated by the blue circle. With this target identified on the ground on Sol N, its position was uplinked on Sol $N + 1$ and a LIBS measurement was taken. Unfortunately, the laser missed the target, so a correction in pointing was made on Earth on Sol $N + 1$ and uplinked on the next opportunity. The ground-in-the-loop pointing refinement was repeated *four times* over a 40-sol period from Sols 185–226 (an anomaly on Sol 200 limited rover operations for for a few weeks thereafter).[1] On the fourth try, a measurement of McGrath was successfully made and was important for understanding the geochemistry of the region. McGrath's analysis was featured in several scientific publications [Léveillé *et al.*, 2014; McLennan *et al.*, 2014].

The lower half of Figure 2 shows an overview of the proposed precision targeting strategy. The process begins in the same way, with an RMI image acquired on Sol N and a target identified within that image (indicated by the blue circle). However, in addition to the estimated target position in the rover coordinate frame, a set of "visual landmarks" are also uplinked to the rover. The visual landmarks are defined by pixel coordinates (in the Sol N image) and visual features that describe properties of the image at those locations. We discuss the choice of features in more detail below.

[1] Some details of planning for each sol can be found under the Science Operations Working Group (SOWG) documentarian notes in the MSL Analyst's Notebook [Stein, 2014].

On Sol $N+1$, the rover points the RMI at the best-guess location of the target and acquires an image. As above, this initial pointing would not hit the intended target. The rover then extracts visual landmarks from this new image and matches them with the landmarks from the Sol N image. This allows the rover to map the target defined in the Sol N image into the Sol $N + 1$ image, and infer that the current pointing will not hit the intended target. Hence, the rover adjusts the pointing accordingly to take a measurement of the intended target, returning the data on Sol $N + 1$, rather than after several ground-in-the-loop attempts to adjust the pointing. This saves at least one sol, or up to 4 attempts over 40 sols as for McGrath, and enables other measurements that might be otherwise unachievable due to time constraints or ongoing thermal shifting between each command cycle.

3 Related Work

There are several constraints on precision instrument targeting to make it practical for rover operations. First, because there is limited storage on board the rover, the Sol N RMI image might be deleted after it is sent to Earth, and not available during subsequent sols. Thus, the visual description of the intended target must be compact so that it can be uplinked to the rover using limited bandwidth. Secondly, because visual targeting must be performed using the relatively limited on-board rover processor (a $133\,\mathrm{MHz}$ RAD750), the visual feature extraction and matching algorithms must be efficient. To avoid adverse impacts on the rover schedule, it is desirable that the entire landmark matching process complete as fast as possible, and run no longer than 10 minutes in the worst case. Finally, because the RMI has such a narrow field-of-view, the algorithm must be able to handle cases when the desired target is not present in the Sol $N + 1$ field-of-view (although there must be some overlap between the images).

Considering these requirements, we can describe the applicability of prior related work. A similar capability already employed on Mars exploration rovers is visual target tracking (VTT). When navigating to a target of interest, rovers use VTT to ensure that they are approaching the correct position over long autonomous drives [Kim *et al.*, 2008]. VTT uses normalized cross-correlation (NCC) to find the most likely location of a 21-by-21 pixel template sub-image containing the target in a wide field-of-view navigation camera image. Commanding VTT requires a small uplink bandwidth and can be efficiently performed using rover hardware. However, the use of NCC with a template image requires that the target of interest actually be in the field-of-view of the follow-up (Sol $N + 1$) image. While this is reasonable for the wide field-of-view navigation cameras, this might not be the case for RMI images. Hence, this approach (or any other NCC-based approach) does not fulfill all of the requirements.

An alternative strategy is to register the Sol N and Sol $N + 1$ images. Given the alignment of two images, the position of the target with respect to the Sol $N + 1$ pointing can be inferred, even if it is not present in the field-of-view. State-of-the-art image registration approaches use visual landmarks detected and matched across a pair of images. Landmarks are locations in an image with associated visual features used to

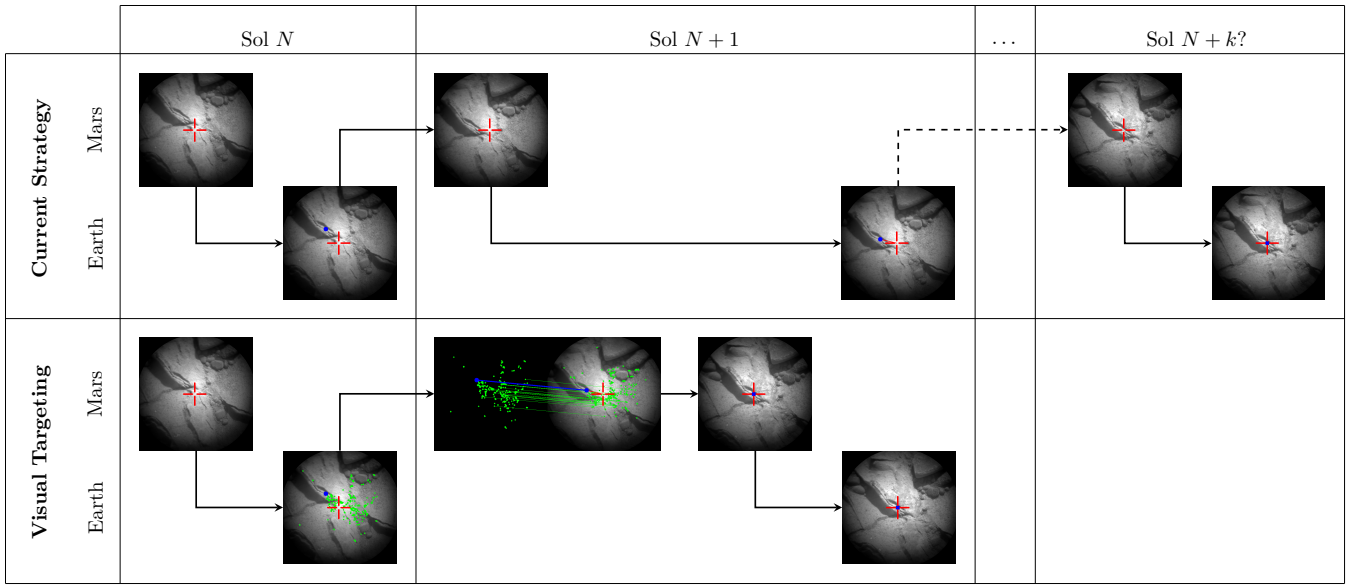

Figure 2: A comparison of precision targeting versus the current strategy. In both cases, an image is acquired on Mars on Sol N and downlinked to Earth, where a target (blue circle) is identified. Currently, the best estimate of the rover-relative target location is used to acquire a measurement on Sol $N+1$. If the target is missed, a correction to pointing is made and uplinked on the next opportunity. This process continues for k sols until the target is measured, or until the rover drives to another location to accomplish other science objectives. In contrast, visual targeting extracts and uploads visual features along with the target location, which are matched with visual features extracted from the Sol $N+1$ image to correct pointing on board. The rover acquires the correct measurement on Sol $N+1$, saving at least one sol or enabling otherwise unachievable measurements.

uniquely describe image the at that location [Moravec, 1981]. Landmarks are matched across two images by finding the nearest-neighbors of these descriptors in the feature space.

A successful and widely used landmark description is scale-invariant feature transform (SIFT), which uses a histogram of image gradient orientations in a region surrounding the landmark [Lowe, 2004]. SIFT has been used for visual targeting with terrestrial robots [Thompson *et al.*, 2011]. Unfortunately, SIFT feature descriptors are large (128-dimensional) and expensive to uplink, and not efficient to compute on board. Even subsequently developed landmark features like speeded up robust features (SURF) [Bay *et al.*, 2006], aimed to reduce the computational complexity of SIFT, are too expensive to compute with rover hardware.

More recently, with an increasing interest in performing image registration for real-time applications using embedded hardware and mobile devices, even more efficient image descriptors such as oriented FAST and rotated BRIEF (ORB) and binary robust invariant scalable keypoints (BRISK) were developed [Rublee *et al.*, 2011; Leutenegger *et al.*, 2011]. In addition to being efficient to compute, these descriptors use binary features, which are more compact to represent than those of SIFT or SURF. A comparison of these binary feature descriptors found that ORB was more efficient than BRISK in terms of memory usage and computational cost, with similar performance [Heinly *et al.*, 2012]. Hence, ORB best satisfies the requirements and was selected for our application.

4 Algorithm

In this section, we describe some details of the pointing refinement implementation (a schematic overview is given in Figure 3). As its name indicates, the implementation of ORB descriptors involves two key components: (1) the detection of landmarks using the features from accelerated segment test (FAST) detector, and (2) the computation of descriptors using binary robust independent elementary features (BRIEF). The FAST detector uses a pre-trained decision tree to detect corners in an image, then uses these corners to orient the landmark (making them invariant to rotation between the two images) [Rosten *et al.*, 2010]. The BRIEF descriptor is a binary string, with each bit representing a comparison of pixel intensities for 256 randomly selected pairs of points within a window surrounding the landmark [Calonder *et al.*, 2010]. Each of these operations involves only simple numerical comparisons that can be implemented efficiently with most processor architectures. Accordingly, ORB is at least 1–2 orders of magnitude faster than SIFT or SURF, with similar performance on the registration task [Rublee *et al.*, 2011]. The 256 binary features can be represented as a 32-byte bit string.

Landmarks are extracted from the Sol N image on Earth and uplinked to Mars. After the Sol $N+1$ image is acquired and landmarks are extracted, the landmarks are matched across the two images. This step involves computing the Hamming distance between all 32-byte feature descriptors and finding nearest neighbors. "Cross-checking" is used so that landmark matches are only accepted if the landmarks in the pair are mutual nearest neighbors to each other.

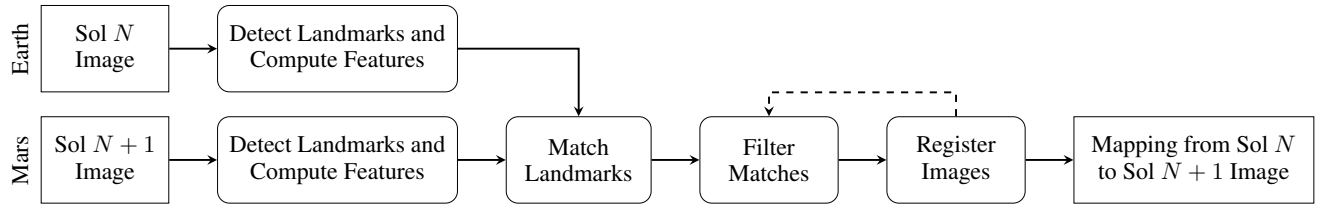

Figure 3: An overview of the pointing refinement algorithm. The steps in the top half of the figure are performed on Earth, while those in the lower half are processed on board the Mars rover.

Although BRIEF features are designed to uniquely describe the region of the image surrounding a landmark, there are often false positive matches between landmarks across the two images. In fact, we observe that for images in our domain, often fewer than 25% of the landmark matches found between images are valid. Hence, the matches must be filtered to remove these false positives. This presents a significant challenge for which we propose a novel solution below.

Assuming that we have filtered out false positive landmark matches, we can register the two images using a homography transformation, which is a sufficient approximation since the rover will not have driven between the Sol N and Sol $N+1$ images and the camera motion is a rotation (though the camera is slightly offset from the axis of rotation). A homography is defined by a 3-by-3 matrix $\mathbf{H}$, where for any $\mathbf{x} = (x, y)$ in the Sol N image pixel space, the corresponding point $\mathbf{u} = (u, v)$ in the Sol $N+1$ image pixel space is given by: $\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \alpha \mathbf{H} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$, for some scale factor α. Decomposing $\mathbf{H}$ as $\begin{bmatrix} \mathbf{A} & \mathbf{t} \\ \mathbf{h}^\top & 1 \end{bmatrix}$, we can represent this relationship in closed form: $\mathbf{u} = h(\mathbf{x}) = (\mathbf{A}\mathbf{x} + \mathbf{t}) / (\mathbf{h}^\top \mathbf{x} + 1)$. We can solve for this 8-degrees-of-freedom homography matrix by using least-squares regression and the known matching landmarks. The solution produces a mapping from the Sol N image to the Sol $N+1$ image, allows the position of the target to be identified with respect to the current pointing, and reveals if it is necessary to adjust the instrument pointing.

To return to addressing the problem of false matches, random sample consensus (RANSAC) is traditionally used for this purpose while simultaneously finding the homography transformation (hence the dashed feedback loop between these two steps illustrated in Figure 3) [Fischler and Bolles, 1981]. RANSAC works by selecting from the matches a random subset of the minimum size required to fit a model (four for a homography). Then, the remaining landmarks are mapped into the new image using this homography transformation, and those that end up close to their corresponding matching landmarks are considered inliers. This process is repeated until a solution with a large fraction of inliers is found.

Unfortunately, as described above, fewer than 25% of landmark matches are inliers in many cases. Thus, the probability that all four samples selected by RANSAC are inliers in a given iteration is less than $0.25^4 \leq 0.4\%$. To be 99% sure that the correct solution is found, over $1,000$ RANSAC iterations would be required, which is computationally infeasible.

We develop a novel strategy for inlier detection using the fact that because the RMI field-of-view is small in comparison to the target distance, the relationship between two images is roughly a translation. That is, in the decomposition of $\mathbf{H}$, $\mathbf{A} \approx \mathbf{I}$ (the identity matrix), and $\mathbf{h} \approx \mathbf{0}$. Our approach is based on the intuition that if we take matching landmark location pairs $(\mathbf{l}_a, \mathbf{l}_b)$ and compute the difference $\mathbf{l}_b - \mathbf{l}_a$, then differences for inliers will all occupy a small region in the two-dimensional pixel space, since the translation between the two landmarks should be roughly the same for all true matches. On the other hand, these differences for false positive matches will be spread out more uniformly. Thus, all we must do is find a region with a high density of points $\mathbf{l}_b - \mathbf{l}_a$, and this region will contain the inliers (the valid landmark matches). This can be accomplished using the density-based spatial clustering of applications with noise (DBSCAN) algorithm [Ester *et al.*, 1996]. Now, we formalize this intuition:

Proposition 1. *Given an image pair related by homography* $\mathbf{H}$ *with inlier landmark density* ρ *(landmarks per pixel), there exists a region where density of landmark difference points* $\mathbf{l}_b - \mathbf{l}_a$ *is at least:*

$$\rho \left(1 - \sqrt{2} s \left\|\mathbf{h}\right\|\right)^2 \Big/ \left(\left\|\mathbf{A} - \mathbf{I}\right\|_2 + 2\sqrt{2} s \left\|\mathbf{h}\right\|\right)^2,$$

where $s = \max(\text{width}, \text{height})$ *of the image (in pixels).*

Proof. For any landmark $\mathbf{x}$, consider all other landmarks in a sphere of radius ϵ centered around $\mathbf{x}$. Then each of these landmarks can be written as $\mathbf{x} + \boldsymbol{\delta}$ for $\boldsymbol{\delta}$ with $\|\boldsymbol{\delta}\| \leq \epsilon$. We will show that the landmark differences $[h(\mathbf{x}) - \mathbf{x}]$ and $[h(\mathbf{x} + \boldsymbol{\delta}) - (\mathbf{x} + \boldsymbol{\delta})]$ are all in a sphere of ϵ', so the density of these difference points is $\rho' = \rho \frac{\pi \epsilon^2}{\pi \epsilon'^2} = \rho \frac{\epsilon^2}{\epsilon'^2}$. Now, we will bound ϵ' by bounding:

$$\begin{aligned}
\|\boldsymbol{\Delta}\| &= \|(h(\mathbf{x} + \boldsymbol{\delta}) - (\mathbf{x} + \boldsymbol{\delta})) - (h(\mathbf{x}) - \mathbf{x})\| \\
&= \left\| \frac{\mathbf{A}(\mathbf{x} + \boldsymbol{\delta}) + \mathbf{t}}{\mathbf{h}^\top (\mathbf{x} + \boldsymbol{\delta}) + 1} - \boldsymbol{\delta} - \frac{\mathbf{A}\mathbf{x} + \mathbf{t}}{\mathbf{h}^\top \mathbf{x} + 1} \right\| \\
&= \frac{\left\| (\mathbf{A} - \mathbf{I}) \boldsymbol{\delta} - \mathbf{h}^\top (\mathbf{x} + \boldsymbol{\delta}) \boldsymbol{\delta} - \mathbf{h}^\top \boldsymbol{\delta} \frac{\mathbf{A}\mathbf{x} + \mathbf{t}}{\mathbf{h}^\top \mathbf{x} + 1} \right\|}{\mathbf{h}^\top (\mathbf{x} + \boldsymbol{\delta}) + 1} \\
&\leq \frac{\|(\mathbf{A} - \mathbf{I}) \boldsymbol{\delta}\| + \left|\mathbf{h}^\top (\mathbf{x} + \boldsymbol{\delta})\right| \|\boldsymbol{\delta}\| + \left|\mathbf{h}^\top \boldsymbol{\delta}\right| \|h(\mathbf{x})\|}{\mathbf{h}^\top (\mathbf{x} + \boldsymbol{\delta}) + 1} \\
&\leq \frac{\|(\mathbf{A} - \mathbf{I}) \boldsymbol{\delta}\| + \|\mathbf{h}\| \|\mathbf{x} + \boldsymbol{\delta}\| \|\boldsymbol{\delta}\| + \|\mathbf{h}\| \|\boldsymbol{\delta}\| \|h(\mathbf{x})\|}{\mathbf{h}^\top (\mathbf{x} + \boldsymbol{\delta}) + 1} \\
&= \frac{\|(\mathbf{A} - \mathbf{I}) \boldsymbol{\delta}\| + \|\mathbf{h}\| \|\boldsymbol{\delta}\| (\|\mathbf{x} + \boldsymbol{\delta}\| + \|h(\mathbf{x})\|)}{\mathbf{h}^\top (\mathbf{x} + \boldsymbol{\delta}) + 1},
\end{aligned}$$

where the first inequality follows from the triangle inequality, and the second follows from the Cauchy–Schwartz inequality. Then, by the definition of the spectral norm $\|(\mathbf{A}-\mathbf{I})\,\boldsymbol{\delta}\| \leq \|\mathbf{A}-\mathbf{I}\|_2 \|\boldsymbol{\delta}\|$. Using this along with Cauchy–Schwartz in the denominator and that $-\left|\mathbf{h}^\top(\mathbf{x}+\boldsymbol{\delta})\right| \leq \mathbf{h}^\top(\mathbf{x}+\boldsymbol{\delta})$:

$$\|\boldsymbol{\Delta}\| \leq \frac{\|\mathbf{A}-\mathbf{I}\|_2 \|\boldsymbol{\delta}\| + \|\mathbf{h}\| \|\boldsymbol{\delta}\| (\|\mathbf{x}+\boldsymbol{\delta}\| + \|h(\mathbf{x})\|)}{1-\|\mathbf{h}\| \|\mathbf{x}+\boldsymbol{\delta}\|}$$
$$\leq \epsilon \frac{\|\mathbf{A}-\mathbf{I}\|_2 + \|\mathbf{h}\| (\|\mathbf{x}+\boldsymbol{\delta}\| + \|h(\mathbf{x})\|)}{1-\|\mathbf{h}\| \|\mathbf{x}+\boldsymbol{\delta}\|}.$$

Now, observe that $\mathbf{x}$, $\mathbf{x}+\boldsymbol{\delta}$, and $h(\mathbf{x})$ are all landmarks in one of the two images, and therefore must be contained in the bounds of the images. So their norms are at most $\sqrt{2}s$:

$$\|\boldsymbol{\Delta}\| \leq \epsilon \left(\|\mathbf{A}-\mathbf{I}\|_2 + 2\sqrt{2}s\,\|\mathbf{h}\|\right) \Big/ \left(1-\sqrt{2}s\,\|\mathbf{h}\|\right).$$

Finally, ϵ' is bounded by $\|\boldsymbol{\Delta}\|$, so substituting this inequality into the expression for the density of landmark difference points, $\rho\frac{\epsilon^2}{\epsilon'^2}$, gives the inequality in the proposition. □

Intuitively, Proposition 1 tells us that as $\mathbf{A} \to \mathbf{I}$ and $\mathbf{h} \to \mathbf{0}$, the density of landmark difference points becomes increasingly large. In fact, in the limiting case when the images are related only by a translation, the landmark difference points are all co-located and thus occupy a region of infinite density. For comparison, differences of randomly matched landmark points (false positives) will have a density of at most ρ, so the region of true matches will have a much higher density than this when $\|\mathbf{A}-\mathbf{I}\|_2$ and $\|\mathbf{h}\|$ are small. Proposition 1 also shows that while the assumption of a near-translation might be insufficient for precise image registration, it can be sufficient for the purposes of filtering outliers.

To find the high-density region of landmark match differences, DBSCAN uses two parameters: a neighborhood size ϵ and minimum number of neighbors m for a point to be an inlier [Ester *et al.*, 1996]. Rather than pick fixed parameters, we use a schedule of increasingly permissible parameters: $(\epsilon, m) = \{(6,8), (7,7), (8,6), (9,5), (10,5), (11,5)\}$. We iteratively run DBSCAN with the parameters until a high-density region is found and then stop. We terminate unsuccessfully if no high-density region is found after all parameter sets are explored. Upon successful termination, we exclude all outlier matches as determined by DBSCAN and use RANSAC to find the final homography between images.

Finally, because spacecraft operations are extremely risk averse, we must build in safeguards to mitigate any errors made during image registration. Even if the registration is successful 99% of the time, we must ensure that the 1% failure cases will not lead to a re-pointing that will cause harm to the spacecraft. A failure can occur for several reasons; for example, there might not be any overlap between the Sol N and $N+1$ images, or there might be too few matching landmarks between the images if the region of overlap is small. Thus, we employ a classifier to automatically detect when a registration failure has occurred.

To evaluate the homography $\mathbf{H}$ produced by our algorithm, we use several features. First, guided by Proposition 1, we use $\|\mathbf{A}-\mathbf{I}\|_2$ and $\|\mathbf{h}\|$ as features. We also use the translations $\mathbf{t}_x$ and $\mathbf{t}_y$ as features, since these should be within the range of the images sizes. Finally, we use ϵ and m parameters selected for DBSCAN and the number of landmark pairs ultimately used to produce the homography matrix as features. Given these seven features, we train a random forest classifier [Breiman, 2001] to evaluate whether the registration was successful. We describe the training and evaluation of this classifier in the following section.

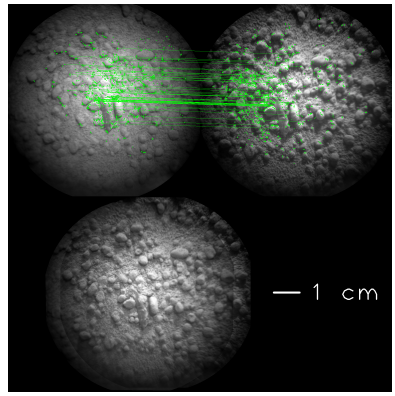

Figure 4: An example of successful registration under significant illumination changes. A merged image is at the bottom. The appearance of the textured surface varies with illumination angle. There are three Mars hours of difference between the images.

5 Evaluation

To evaluate the ability of our proposed precision targeting approach to identify targets in an RMI image and to characterize the factors that affect its performance, we used over 3,800 images acquired by the MSL rover during the first 1,200 sols of operation.[2] Using metadata that includes the instrument pointing, we identified over 16,500 pairs of images that should overlap by less than the $19\,\mathrm{mrad}$ field-of-view. We are interested in evaluating how two factors affect performance: (1) initial pointing accuracy and (2) changes in illumination conditions between the Sol N and Sol $N+1$ images (as in Figure 4). First, we describe several implementation details.

To implement the precision targeting algorithm described above, we used the OpenCV implementation of ORB feature extraction and RANSAC-based homography finding [Bradski, 2000; Rublee *et al.*, 2011]. A random forest with 100 trees was used to classify the homographies produced by the process in Figure 3. To train the random forest classifier, we analyzed the roughly 16,500 alignments produced by our algorithm using 2,500 landmarks. We used the noisy "ground truth" RMI metadata to identify obviously successful cases when more than 150 landmark pairs made it past the filters and the offset was within $2\,\mathrm{mrad}$ of the reported pointing discrepancy between the two images. Similarly, we selected cases for which less than 10 landmark pairs were found and the estimated discrepancy was more than $20\,\mathrm{mrad}$ off from what was reported. Other cases were ambiguous due to small errors in reported pointing, so these were labeled manually.

The resulting classifier has a 97% true-positive rate at a 0.5% false-positive rate (using the properly-selected threshold) as estimated using the labels on the training examples when they are left out of bootstrap samples. The $\|\mathbf{A}-\mathbf{I}\|_2$, $\|\mathbf{h}\|$, and number of landmark matches after filtering turn out

[2] Images are made publicly accessible online via the Planetary Data System [Wiens, 2013].

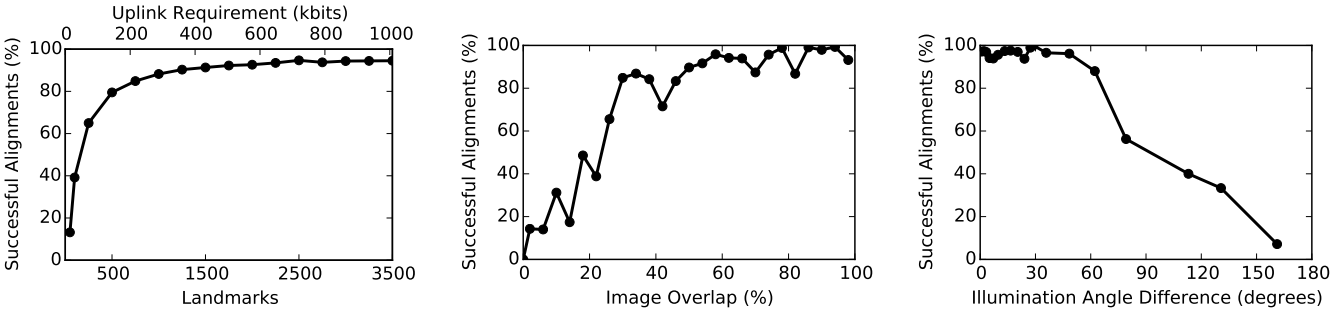

Figure 5: Performance as a function of three factors: (**left**) the number of landmarks initially extracted from the images (and therefore the amount of information that must be uplinked to the spacecraft), (**center**) the amount of overlap between the Sol N and $N+1$ images, and (**right**) the difference in illumination angle between the images.

to be the most important features as determined by their frequency of use within the random forest. In fact, over 96% of the manually identified correct alignments are such that the density of landmark differences are greater than the background density of randomly matched landmark differences as determined by the expression in Proposition 1. This observation justifies the use of DBSCAN to filter matches.

Given the trained classifier, we can now determine the estimated success rate when different numbers of landmarks are used, as shown in Figure 5 (left). The algorithm was rerun with different numbers of landmarks ranging from 50 to 3,500. The corresponding uplink requirement (in kilobits) is illustrated along the top of the figure. These results show that good performance can be achieved with a reasonable uplink budget of several hundred kilobits. The remainder of the results use 2,500 landmarks, where the pointing refinement approaches its asymptotic performance.

To evaluate how initial pointing accuracy and lighting conditions affect performance, we use two measurements available via the image metadata. The initial pointing accuracy is related to how much overlap there is in the fields of view of the two images, which is related to how much the pointing differs between these images. Let β be the angular difference in pointing between the two images expressed as a fraction of the instrument field-of-view ($19\,\mathrm{mrad}$). Then the area of overlap, as a fraction of the circular field-of-view area, is: $f_{\mathrm{overlap}} = \frac{2}{\pi}\left[\arccos(\beta) - \beta\sqrt{1-\beta^2}\right]$. Figure 5 (center) shows the registration performance as a function of this quantity. The success rate of the approach is stable above 80% after there is roughly 33% overlap between the images.

Finally, we evaluate the effects of changing illumination conditions between images. The solar illumination azimuth and elevation angles at the time the image was taken are also included in the image metadata. The average success rate binned by illumination angle difference is shown in Figure 5 (right). The results suggest that the algorithm is sensitive to illumination changes exceeding $60\,\mathrm{deg}$, likely because shadows formed by textured rock surfaces can vary drastically with lighting angle (see Figure 4), making it challenging to find corresponding landmarks. However, this challenge is mitigated by the fact science operations on Mars typically occur within a consistent block of time each sol, when illumination conditions are similar.

Finally, we have implemented a `C++` version of our precision targeting algorithm, with running times of less than 1 minute with over 3,500 landmarks on a $150\,\mathrm{MHz}$ LEON4 processor, which is comparable to the processing capability of the rover's on-board processor. Thus, the results indicate that our approach satisfies the requirements for operation.

6 Conclusion

In this paper, we present an image-registration-based approach for precise instrument targeting on board the Mars 2020 rover. We demonstrate using real data from the MSL rover that given consistent lighting conditions (within 1–2 Mars hours) and at least 33% overlap between images, our method can correctly identify the target location specified in the first image within the second image in over 80% of cases. Each of these 80% success cases translates to time saved or a new capability to target small features relevant to astrobiology and sample return selection. For the remaining cases, we employ a classifier to reject failed image alignments and fail safe by falling back to the default strategy of measuring the target at the best-guess location coordinates.

Though the analysis in this work is targeted for the SuperCam instrument, there are other instruments selected for the Mars 2020 rover that would benefit from such technology. For example, the Planetary Instrument for X-Ray Lithochemistry (PIXL) [Allwood *et al.*, 2014] and the Scanning Habitable Environments with Raman & Luminescence for Organics & Chemicals (SHERLOC) [Beegle *et al.*, 2015] instrument have small fields-of-view and are located on the rover's arm, which is also subject to thermal and mechanical errors in placement. In addition to investigating the use of pointing refinement for other instruments like PIXL, we are exploring cross-camera pointing refinement, so that the image acquired from one instrument or engineering camera can be used to refine the pointing of another instrument, reducing the downlink requirement and further increasing the efficiency of making valuable scientific measurements.

Acknowledgments

We would like to thank Raymond Francis, Daniel Gaines, Vandi Verma, and Roger Wiens for feedback they provided during the development of this work. This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. Copyright 2016 California Institute of Technology. U.S. Government sponsorship acknowledged.

References

[Allwood *et al.*, 2014] A. Allwood, J. Hurowitz, L. Wade, R. Hodyss, and D. Flannery. Seeking ancient microbial biosignatures with PIXL on Mars 2020. In *American Geophysical Union Fall Meeting Abstracts*, volume 1, 2014.

[Bay *et al.*, 2006] H. Bay, T. Tuytelaars, and L. van Gool. SURF: Speeded up robust features. In *Proceedings of the European Conference on Computer Vision*, 2006.

[Beegle *et al.*, 2015] L. Beegle, R. Bhartia, M. White, L. DeFlores, W. Abbey, Y.-H. Wu, B. Cameron, J. Moore, M. Fries, A. Burton, et al. SHERLOC: Scanning habitable environments with Raman & luminescence for organics & chemicals. In *Proceedings of the IEEE Aerospace Conference*. IEEE, 2015.

[Bradski, 2000] G. Bradski. The OpenCV library. *Dr. Dobb's Journal of Software Tools*, 2000.

[Breiman, 2001] L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[Calonder *et al.*, 2010] M. Calonder, V. Lepetit, C. Strecha, and P. Fua. BRIEF: Binary robust independent elementary features. In *Proceedings of the European Conference on Computer Vision*, pages 778–792, 2010.

[Ester *et al.*, 1996] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the International Conference on Knowledge Discovery and Data Mining*, 1996.

[Fischler and Bolles, 1981] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the Association for Computing Machinery*, 1981.

[Heinly *et al.*, 2012] J. Heinly, E. Dunn, and J.-M. Frahm. Comparative evaluation of binary features. In *Proceedings of the European Conference on Computer Vision*, 2012.

[Kim *et al.*, 2008] W. S. Kim, J. J. Biesiadecki, and K. S. Ali. Visual target tracking on the mars exploration rovers. In *Proceedings of the 9th International Symposium on Artificial Intelligence, Robotics, and Automation in Space*, 2008.

[Leutenegger *et al.*, 2011] S. Leutenegger, M. Chli, and R. Y. Siegwart. BRISK: Binary robust invariant scalable keypoints. In *Proceedings of the IEEE International Conference on Computer Vision*, 2011.

[Léveillé *et al.*, 2014] R. J. Léveillé, J. Bridges, R. C. Wiens, N. Mangold, A. Cousin, N. Lanza, O. Forni, A. Ollila, J. Grotzinger, S. Clegg, et al. Chemistry of fracture-filling raised ridges in Yellowknife Bay, Gale Crater: Window into past aqueous activity and habitability on Mars. *Journal of Geophysical Research: Planets*, 119(11):2398–2415, 2014.

[Lowe, 2004] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 2004.

[Maurice *et al.*, 2012] S. Maurice, R. C. Wiens, M. Saccoccio, B. Barraclough, O. Gasnault, O. Forni, N. Mangold, D. Baratoux, S. Bender, G. Berger, et al. The ChemCam instrument suite on the Mars Science Laboratory (MSL) rover: science objectives and mast unit description. *Space Science Reviews*, 170(1-4):95–166, 2012.

[Maurice *et al.*, 2015] S. Maurice, R. C. Wiens, R. Anderson, O. Beyssac, L. Bonal, S. Clegg, L. DeFlores, G. Dromart, W. Fischer, O. Forni, et al. Science objectives of the supercam instrument for the mars2020 rover. In *Lunar and Planetary Science Conference*, volume 46, page 2818, 2015.

[McLennan *et al.*, 2014] S. McLennan, R. Anderson, J. Bell, J. Bridges, F. Calef, J. Campbell, B. Clark, S. Clegg, P. Conrad, A. Cousin, et al. Elemental geochemistry of sedimentary rocks at Yellowknife Bay, Gale crater, Mars. *Science*, 343(6169):1244734, 2014.

[Moravec, 1981] H. P. Moravec. Rover visual obstacle avoidance. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 785–790, 1981.

[Rosten *et al.*, 2010] E. Rosten, R. Porter, and T. Drummond. Faster and better: A machine learning approach to corner detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(1):105–119, 2010.

[Rublee *et al.*, 2011] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski. ORB: an efficient alternative to SIFT or SURF. In *Proceedings of the IEEE International Conference on Computer Vision*, 2011.

[Stein, 2014] T. Stein. The MSL analyst's notebook. `https://an.rsl.wustl.edu/msl/mslbrowser/`, 2014. NASA Planetary Data System.

[Thompson *et al.*, 2011] D. R. Thompson, D. S. Wettergreen, and F. J. C. Peralta. Autonomous science during large-scale robotic survey. *Journal of Field Robotics*, 28(4):542–564, 2011.

[Wiens *et al.*, 2012] R. C. Wiens, S. Maurice, B. Barraclough, M. Saccoccio, W. C. Barkley, J. F. Bell III, S. Bender, J. Bernardin, D. Blaney, J. Blank, et al. The ChemCam instrument suite on the Mars Science Laboratory (MSL) rover: Body unit and combined system tests. *Space Science Reviews*, 170:167–227, 2012.

[Wiens, 2013] R. C. Wiens. MSL ChemCam remote micro-imaging camera raw data. `https://pds.jpl.nasa.gov/ds-view/pds/viewDataset.jsp?dsid=MSL-M-CHEMCAM-RMI-2-EDR-V1.0`, 2013. NASA Planetary Data System.