Improved Target Handoff for Single Cycle Instrument Placement

Dr. Richard Madison
Jet Propulsion Laboratory, California Institute of Technology
M/S 125-109, 4800 Oak Grove Dr
Pasadena, CA 91109
818-354-1897

Abstract—Single Cycle Instrument Placement[1, 2] (SCIP) could dramatically increase the speed of various planetary rover operations. JPL is validating SCIP for use on its upcoming Mars Science Laboratory mission. Two major sources of error in the implementation submitted for validation were imprecision in selecting a distant target and error introduced while handing the target off to the rover's hazard cameras (hazcams). We have added the capability to designate a target using high resolution cameras (pancams) and then hand off to medium resolution cameras (navcams) with little error later in the traverse. We implemented several options for handoff from navcams to hazcams. We evaluated their performance on several test cases. Two methods produced average handoff error of 1 pixel when successful, but these only worked in half of the tests. Two other methods reliably produced about 10 pixels of error. None reliably reach the goal of 1 pixel handoff accuracy.

TABLE OF CONTENTS

INTRODUCTION

Robotic activity on Mars is not amenable to teleoperation from Earth, due to the long communications delay between the two planets. Instead, operators transmit commands during one command cycle, and a robot follows the commands and reports results in the following cycle. The number of cycles required for an operation depends on the amount of autonomy granted to the robot. Instrument placement typically requires three cycles: one to choose a target and order the robot to approach to a safe distance while avoiding obstacles; a second to reacquire the target in workspace cameras and order the robot to move within arm's reach of the target; and a third to pinpoint the target and order the robot to servo its instrument arm to the target. The robot could perform Single Cycle Instrument Placement (SCIP) if it could accurately track a target from initial designation, through approach and camera handoff, to pinpoint the target for the instrument arm. This would triple the speed of instrument placement operations as well as similar pick-and-place operations such as sample acquisition, resource collection and return to a processing facility, and construction.

A SCIP Implementation

Variations of SCIP have been and continue to be developed and demonstrated, including several recent examples [1] [2] [3] [4]. The last improved components from several of these efforts and began modularizing the framework to simplify the process of swapping out individual components. This version of SCIP is now being validated by the Mars Technology Program at JPL as a step toward integrating the technology into MSL (Mars Science Laboratory) and other future Mars missions. We have continued to enhance and modularize this version of SCIP to support the validation testing. The high-level algorithm of this version of SCIP, running on the Rocky8 rover, consists of the following steps.

(1) The rover generates a panorama of images using its medium field-of-view, mast-mounted, navcams. A user designates a 2D target in one image. SCIP uses stereo processing to identify the target's 3D location.

(2) SCIP uses an obstacle-avoiding navigation algorithm to drive to a point approximately 2m in front of the target.

(3) The navigation algorithm stops the rover frequently to detect obstacles. At each stop, SCIP points the navcams toward the 3D target, takes images, tracks the 2D target into the new images, and uses stereo to update the target's 3D coordinates.

(4) Around 2m from the target, the navcams are too high above the ground to track the target. SCIP turns the rover to point its body-mounted, high field-of-view, hazcams at the target. SCIP then tracks the 2D target into the hazcam images.

(5) SCIP drives to within 1m of the target, stopping several times to take hazcam images, track the target into them, and update the targets 3D coordinates.

(6) About 1m from the target, SCIP uses the rover's manipulator arm to place an instrument on the target.

---

[1] 0-7803-9546-8/06/$20.00© 2006 IEEE
[2] IEEEAC paper #1073, Version 1, Updated Oct, 31 2005

This algorithm tracks the 2D target rather than asking a user to reacquire it at the 2m and 1m stops. The rover's pose estimator accumulates 6-DOF error during the traverse to the target, but this error does not impact the relative 3D position of the target, which is recomputed at each stop after tracking the target into the newest images. In addition, as the cameras approach the target, they are able to provide more accurate stereo ranging. Thus the 3D estimate of the target position actually becomes more accurate over the course of the approach. This scheme only requires that the estimates of target position and rover motion be accurate enough to allow SCIP to point the cameras correctly enough that the target appears in the images.

Pancam-to-Navcam Handoff

Validation testing determined that, at the nominal starting distance of 10m from a target, navcam resolution of about 1cm/pixel is too low to identify interesting targets. To solve this, we began using the narrow field-of-view pancams for target designation and initial tracking. The pancams see about 3mm/pixel at 10m away, allowing more precise target designation. However, they are not useful for tracking the target at close range. SCIP must point the pancams to within about 6° of the target to capture the target in the field of view of the cameras. As the rover approaches a target, error in pose estimation and stereo ranging combine to produce angular pointing errors larger than that, and SCIP risks losing the target. Thus, at 4-5m from the target, SCIP hands off tracking to the navcams, which have three times the field of view and therefore three times the tolerance to pointing error.

On the Rocky8 rover, and by extension on the similarly-configured MER rovers, handoff from pancams to navcams is straightforward. At 4-5m from the target, SCIP takes images with both the pancams and the navcams. After tracking the target into the pancam image, SCIP extracts a window around the 2D target in the left pancam image and shrinks it by the ratio of navcam to pancam fields of view , approximately a factor of one third. The shrunken window reflects the expected appearance of the target in the navcam images. SCIP tracks this new, smaller window into the actual navcam images. This approach works well because the two camera pairs have approximately the same viewpoint, as they all sit in a line on the same pan-tilt platform at the top of the rover's mast. The handoff accuracy has not been validated yet, but initial testing shows that it is as accurate as manual handoff by the user who selects the initial target.

One might argue that, although the pancams have three times the pointing accuracy requirements, they also have three times the stereo ranging accuracy, and so could be pointed with three times the accuracy. Apparently this is not the case. Given a pair of images, the associated camera models, and 2D coordinates of a target in one image, it is possible to calculate an "epipolar" line through the second image, along which the target must lie. Initial testing found that targets in the left pancam image routinely lie up to four pixels off the epipolar line in the right image, suggesting that the camera models are misaligned. This is reasonable given that: the rover took images shortly after stopping, while the chassis may still be rocking as its suspension damps out the deceleration; the cameras sit on a mast, at the end of a long moment arm that amplifies any vibration; and the cameras currently are not synchronized, so they may move between taking the left and right images. We perform stereo matching despite this misalignment by searching for matches in a band around the epipolar line. The same angular misalignment should produce only a one-pixel offset in navcam images (which have 1/3 the angular resolution), and in fact, we found stereo matches when searching the epipolar line in navacam images. However, if the images have similar error offsets along the epipolar line, the resulting four- and one-pixel disparity error would overshadow subpixel-matching error and drive stereo ranging error. Then pancams and navcams would have comparable stereo range error, though pancams would still have three times the pointing requirements.

Navcam-to-Hazcam Handoff

When the rover reaches about 2m from the target, SCIP hands the target off from the navcams to the hazcams. This is necessary because the target is difficult to keep in the navcams' view, both because small uncertainties in rover motion produce large pointing errors and because the navcams simply cannot point that far down. Also, the forward hazcam view of the target is more similar to the originally selected target than is the essentially overhead navcam view, making it easier to track the feature using hazcams.

Unlike pancam-to-navcam handoff, navcams and hazcams do not share a similar viewpoint. Thus, SCIP generally cannot scale a window in the navcam image and track it into the hazcam image. Instead, [4] creates a stereo point cloud from the navcam images and projects the cloud into hazcam image coordinates to create an "expected" hazcam image. This implementation of SCIP then tracks the target from the expected image to the actual hazcam image. Most of the "several pixel" error observed in [4] was attributed to handoff error.

Our preliminary tests found that this straightforward handoff method did not come close to the SCIP goal of one-pixel total error or the 1.6 to 2.7cm total SCIP error reported in [5]. We then began implementing and evaluating alternative methods, mainly derived from those in [13]. This paper describes these methods, explains how we evaluated them, and gives the results of our evaluation.

HANDOFF METHODS

We considered twelve handoff methods. Half of these handoff from "near navcams" while the other half handoff from "far navcams", as depicted in figure 1.

Figure 1 – Near and far navcams.

Near vs. Far Navcams

Half of the methods handoff from *near* navcam images taken at the same rover position as the hazcam images, at the time of handoff. SCIP rotates the rover body right before handoff so that the navcams and hazcams view the target along the same azimuth. The target may still look very different to the two cameras, because the navcams look down on top of the target and the hazcams look down at only 45 degrees. However, we will know the transform between the camera frames, allowing us to project 3D points between them.

The other half of the methods handoff from *far* navcam images taken earlier in the traverse, when the pointable navcams pointed down at the same angle as the fixed hazcams. If the rover approaches the target in a straight line, these navcam images have nearly the same line of sight to the target as the hazcams have at handoff, so a window around the navcam target can be scaled and tracked into the hazcam image, as with pancam-to-navcam handoff. This is fast and avoids the risk of incorrect tracking as the target changes shape in the navcam images, as the rover approaches the target and the navcams get an increasingly overhead view. If the rover approaches the target at an angle, for instance after arcing to avoid obstacles, the viewpoints of the hazcams and the far navcams will differ, so we may not be able to track the feature into the hazcam image. Unlike near hazcam tracking, we probably do not know the transform between the navcam and hazcam frames accurately. Our estimate, derived from our rover pose estimate, includes any pose error accumulated during the navigation.

Pure kinematic handoff (near and far navcam)

The first two methods use stereo on the navcam images to convert a 2D target to a 3D point; transform this point from navcam 3D coordinates to hazcam 3D coordinates; and use the hazcam camera model to project the point to an expected 2D position in the hazcam image

These two methods are called *kinematic* because they rely chiefly on the kinematic chain of the rover mast to provide the 3D transform between navcam and hazcam coordinate systems. This transform is also called the *mast calibration*. For handoff from far navcams, the pose change of the rover after taking the far navcam images is multiplied onto the mast calibration to produce the proper transform.

The 3D target coordinates are generated using point stereo. Point stereo matches the target from the left image by searching a band around the epipolar line in the right image. Searching a band gives some resistance to inaccurate camera models.

Pure kinematic handoff is fast and computationally cheap. Sources of error include incorrect navcam stereo (poor calibration, poor matching, or vibrating mast), incorrect mast kinematic chain (poor calibration or vibrating mast), incorrect hazcam model, and when using far navcams, incorrect pose estimation. Mast vibration is an issue because the cameras are not synchronized, and any mast motion between imaging causes cameras to have different references frames and thus poor calibration.

Alternate pure kinematic handoff (near and far navcam)

The next two methods differ from the first two only in that they use JPL Stereo [6] instead of point stereo to determine 3D coordinates. JPL Stereo is optimized to quickly generate stereo data for an entire image. It uses the camera models to rectify the input images so that pixels in the left image match pixels on the same scan line in the right image. Thus, it relies on accurate camera models and may be more sensitive to poor calibration of navcams.

JPL stereo is slower than point stereo, because it must generate an entire image worth of stereo rather than a single point. However, if SCIP computes such stereo anyway, for instance for navigation or pose estimation, then alternate pure kinematic handoff will be the faster option.

Refined kinematic handoff (near and far navcam)

The next two methods expand on kinematic handoff by projecting not just the navcam target but a window around it into the hazcam model to produce an *expected* hazcam target. They track this expected target into the actual hazcam image using normalized cross correlation. Normalization is important, as the autogain of the cameras can cause significant intensity differences between the navcam and hazcam images. Refined kinematic handoff can be done from near or far hazcam images.

Refined kinematic handoff from the near navcam images uses the following steps to build the expected target.

(1) Use JPL stereo to build a point cloud. Each point takes the intensity of the associated navcam pixel.

(2) Project these points through the hazcam model to generate an expected hazcam image. Where multiple

points project to the same hazcam pixel, accept the point nearest the hazcam. Where a point projects between four hazcam pixels, allow all four to accept the point. Even with such blurring, the expected image may contain holes through which points from the far side of the point cloud may be visible. This may cause errors in the tracking under the current algorithm.

(3) Extract the expected target window from the expected image. Center the window at the alternative pure kinematic handoff point, which by construction is the location of the target in the expected hazcam image.

Refined kinematic handoff from far navcam images extracts a window from the navcam image, scales it, and tracks it as described under pancam-to-navcam handoff. The scale factor consists of the ratios of fields-of-view and distances-to-target for the two camera sets. The search for a comparable window in the hazcam image can be centered about any of the pure kinematic handoff points. We use the one calculated from near navcams and point-stereo, as it is the most accurate.

Handoff from far navcam images is particularly appropriate when the line of sight to the target is the same for hazcams and far navcams. We did not explore the option of warping the expected feature to account for differing line of sight. This could have been done using affine warping or using the same method as in handoff from near hazcams.

In theory, both methods require *rectified* navcam and hazcam images – images that have been warped to undo the effect of radial distortion in the cameras. Then the target appearance would not depend on its location in the image, and so would be easier to track. The methods should *rectify* the images, then generate and track the expected target, then unrectify the result back to the original hazcam model. In practice, this is unnecessary because the regular tracking keeps the target near the image center, where distortion is minimized.

Refinement is slower than pure kinematic handoff, but it compensates for a number of the possible error sources of pure kinematic handoff. Errors in camera pointing will tend to translate the image of the target but not rotate it much, so the expected feature will be more or less accurate, and the tracking will locate it. When using near navcam images, navcam calibration errors can cause a warping of the 3D structure around the target, which can create an incorrect expected feature that NCC cannot track. This often causes NCC to fail spectacularly, though the failure is not always betrayed by a low correlation peak. When using far navcam images, failure to approach the target directly can cause a warping of the 2D expected feature and the same resulting failure in tracking.

Pure geometric handoff (near and far navcam)

The next two methods generate a point cloud from hazcam stereo, project it into the navcam images, and identify the point that projects nearest the 2D target. It projects that point back into the hazcam image as the handoff point. The name, *geometric* handoff, carries no particular significance.

Pure geometric handoff is slower than pure kinematic handoff, because it projects the entire point cloud, not just one point. However, it may be preferable when hazcam stereo is more accurate than navcam stereo. This is generally the case, because navcams are harder to calibrate and so calibrated less often, and because navcams sit at the end of a long lever and possibly moving arm, farther from the target. Geometric handoff can be done from near or far navcams, but projecting into far navcams requires use of the noisy pose estimate.

Refined geometric handoff (near and far navcam)

The next two methods generate and track an expected hazcam target, as in refined kinematic handoff, but they use a hazcam stereo point cloud, as in pure geometric handoff. Each hazcam pixel is projected into the 3D point cloud, converted to navcam 3D coordinates using mast calibration and any rover pose change, then projected into navcam 2D coordinates. The navcam image intensity at those coordinates is copied to the hazcam pixel. Effectively, this "drapes" the navcam image over the surface defined by the hazcam stereo.

Generating the expected target does not take significantly more time than finding the handoff point under pure geometric handoff, though the refinement step takes extra time.

Refined geometric handoff may be better than refined kinematic if hazcam stereo is more accurate than navcam stereo. However, any mast calibration, pose estimation, or hazcam calibration errors will cause the navcam image will drape incorrectly over the hazcam stereo, often distorting the expected feature and making it untrackable. Flat targets on large, flat areas should be robust to this error. Refined geometric handoff from far navcams is likely to fail unless the pose estimate is very accurate or the target is on a large, flat area.

As with refined kinematic handoff, it might more sense to generate the expected feature using a rectified hazcam model, rectify the actual hazcam, track, and then unrectify the resulting handoff point. In practice, if the target is near the center of the image, this should not be necessary.

An option we did not explore is to project the hazcam stereo onto the navcam frame to create a fake navcam image, track the actual navcam target into the fake navcam image, and then project the tracked point back into the hazcam image. That might work for far navcams, but it probably would not work for the near navcams, where the hazcam data is too sparse to fill much of a fake hazcam image.

Mesh registration handoff (near and far navcam)

The remaining two methods do pure kinematic handoff except that instead of using mast calibration or pose estimate to convert between navcam and hazcam 3D coordinates, they independently recover the transform using *mesh registration* [7][8]. Mesh registration is an improvement on the Iterative Closest Point (ICP) [9] algorithm, which finds the transform that best aligns two point clouds. We use JPL stereo to generate point clouds from both navcam and hazcam images, use mesh registration to find the transform between the clouds, and apply kinematic handoff with using that transform. In addition, as explained later, for near navcam handoff, we seed mesh registration with the mast calibration.

Mesh registration uses a Nelder-Mead optimization, making it slow compared to the other methods.

Errors in camera pointing do not impact this handoff method, making it potentially more accurate than the other methods. For instance, [5] achieved 0.5cm-1.5cm handoff error, which corresponds to 2-6 pixels for Rocky8 hazcams 1m from the target. On the other hand, stereo point clouds can be very difficult to align, due to both mesh distortion resulting from matching and camera calibration errors, and the need to segment the point clouds into overlapping and non-overlapping regions. [10] found ICP unable to align meshes generated by JPL Stereo. Perhaps the new mesh registration method will solve these problems.

EXPERIMENT

To evaluate the relative merit of the handoff methods, we performed each of the 12 types of handoff on each of 18 targets. These consisted of three sets of three targets, which we approached in a direct line and along an arc. The images containing these targets are shown in figures 2 and 3.

We placed the Rocky8 rover in front of each of three scenes, with the cameras about 1m (ground distance) from a large rock. We waited for the rover vibrations to damp out and then took images with both (near) navcams (figures 2a and 3a) and hazcams (figures 2c and 3c). We moved the rover straight back until the cameras were about 3m (ground distance) from the same rock, pointed the navcams at the rock, waited for vibrations to damp out, and took (far) navcam images (figures 2b and 3b). At this distance, the far navcams have approximately the same line of sight to the large rock as the hazcams had when we took hazcam images. Next, we repeated the three experiments except this time we moved the rover back along an arc, keeping the navcams pointed at the rock, until the cameras were 3m from the rock and the angle from original position to rock to new position was about 30 degrees, so that the rover's view of the rock has rotated by about 30 degrees.

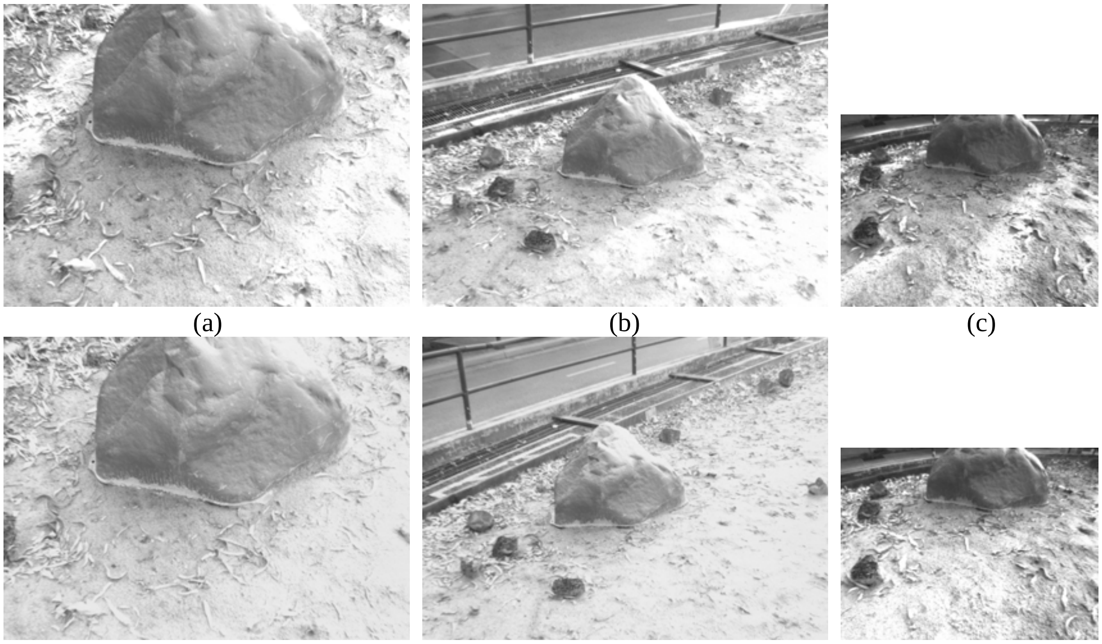
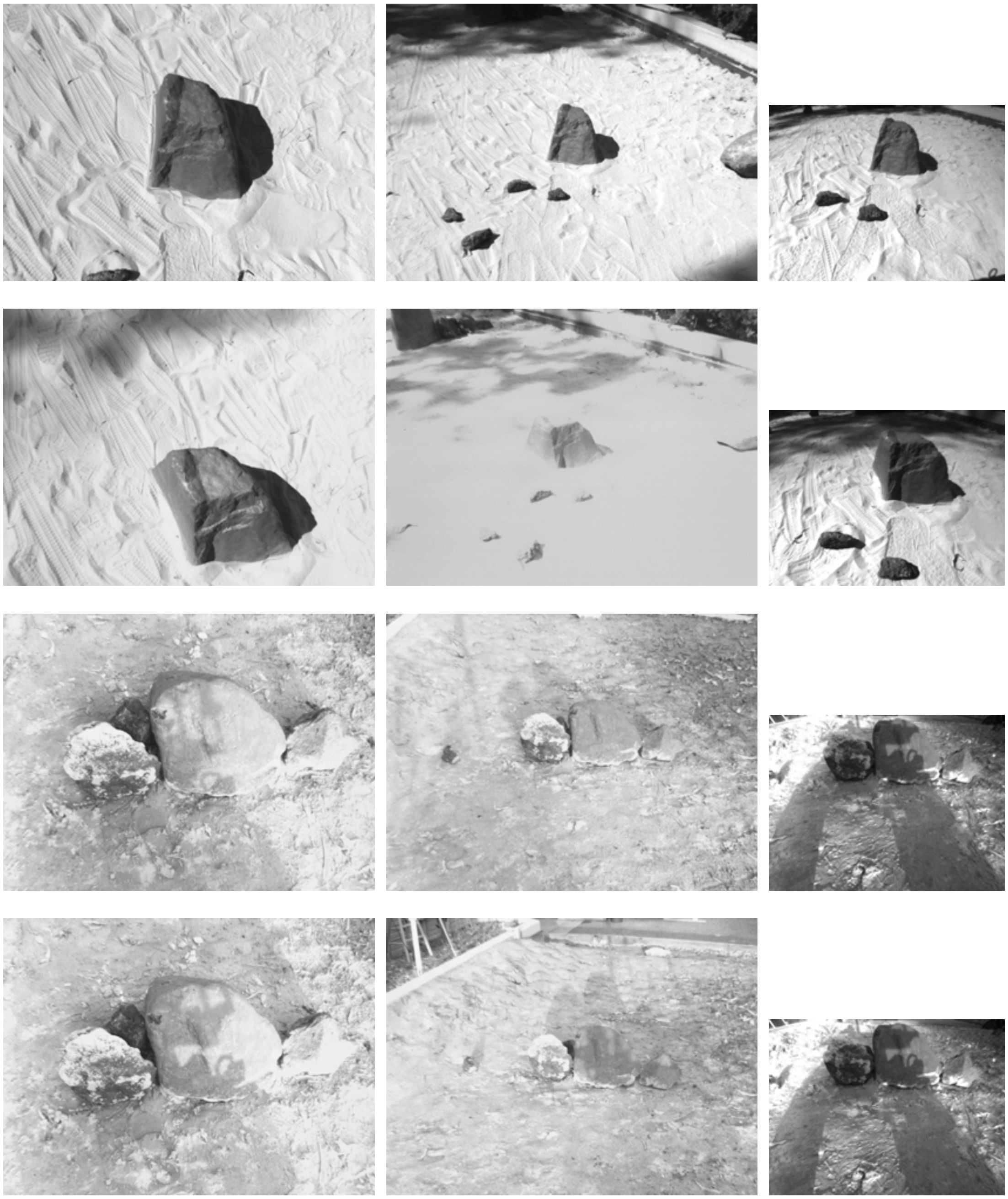

(a) (b) (c)

Figure 2 – Left near navcam (a), far navcam (b), and hazcam (c) images from test runs 1 and 2, two approaches to the same rock. Near navcam and hazcam images are similar because the rover ends at the same place. Far navcam images differ by a 30° rotation, to test the effect of targets rotating out-of-plane.



(a) (b) (c)

Figure 3 - Left near navcam (a), far navcam (b), and hazcam (c) images from test runs 3-6, two pairs of approaches to two rocks. Near navcam and hazcam images are similar because the rover ends at the same place. Far navcam images differ by a 30° rotation, to test the effect of targets rotating out-of-plane. The far navcam image in test 4 is badly washed out, but handoff algorithms must accommodate this common situation.

For each of the six tests, we identified three targets. The goal was to have targets on flat places, on sharp points, occluding contours, sand, and high contrast points. We manually identified the positions of the features in all of the images. We applied each of the 12 handoff methods to convert the navcam target positions into hazcam target positions. The "ground truth" manually established from the hazcam images has error bars of a few pixels. This seems reasonable because we expect the accuracy of the methods to differ by much more than this.

Before applying the handoff methods to the data, we calibrated the hazcams using a standard, JPL, surveyed calibration procedure [12]. The hazcams very clearly had been out of calibration, producing poor stereo recovery and unusually bad geometric handoff in preliminary tests. We did not calibrate either the navcams or the mast pointing. Neither is calibrated regularly, as there is currently no simple technique to calibrate the mast or to accurately identify the reference frame in which navcams are being calibrated. Algorithms operating on Rocky8 must accept some error in both, so the tests were run with the error in place. As a result, the tests will measure handoff performance under conditions normally experienced on Rocky8 rather than at their theoretical best.

Table 1. Summary of Handoff Accuracies

Method	Navcam	Success	Average Error (pixels)	Error Range (pixels)
Pure Kinematic	Near	100%	13 ± 3	9-23
	Far	100%	107 ± 58	63-182
Alternate Pure Kinematic	Near	100%	22 ± 5	15-35
	Far	83%	102 ± 27	68-152
Refined kinematic (success / failure)	Near	55%	1 ± 1 / 99 ± 76	0-3 / 18-214
	Far	67%	2 ± 1 / 83 ± 52	0-4 / 16-166
Pure Geometric	Near	100%	9 ± 6	1-22
	Far	100%	54 ± 21	20-95
Refined geometric (success / failure)	Near	50%	1 ± 1 / 109 ± 77	0-3 / 8-261
	Far	17%	2 ± 1 / 52 ± 55	1-3 / 19-231
Mesh registration	Near	100%	28 ± 11	8-58
	Far	100%	124 ± 57	36-226

RESULTS AND DISCUSSION

Table 1 summarizes the accuracy of the various handoff methods, giving the average error for each method over the 18 test targets, the deviation and range of errors, and the fraction of the targets for which each method succeeded. When they work, refined kinematic and refined geometric methods recover the correct handoff pixel to within measurement error, approximately 2 pixels. However, these methods are not particularly reliable. Among the reliable methods, pure geometric handoff from the near hazcams gives the best results, averaging 9 pixel error, followed by pure kinematic handoff using point stereo, which averages 13 pixel error.

Pure kinematic handoff from near navcam

In this method, 2D target coordinates in the navcam image are projected into 3D using stereo, transformed to rover body frame, and projected into the hazcam using the hazcam model.

Errors averaged 13 pixels, with a large systematic bias. Specifically, the actual target in the hazcam images was consistently about 13 rows below the predicted position. This suggests two potential error sources. Miscalibration of the navcams could produce a consistent underestimate of distance to target. The observed offset would require approximately 5% (7cm) underestimate in target distance. For reference, a recent hazcam recalibration altered depth estimates by 10%. Another explanation is miscalibration of the mast pointing. A 1 degree tilt error in the mast could produce the observed 13 pixel offset. Ongoing and yet-unpublished work at JPL on autonomous mast calibration adjustment recently found a nearly 1 degree tilt error in mast calibration. Mast calibration error is sufficient to explain the observed error, but there is no obvious mechanism by which the mast would go out of calibration by a full degree, whereas we have previously observed cameras going out of calibration enough to cause a 5% ranging error. One possibility is that the mast calibration adjustment corrects for both error sources, at least in the neighborhood of the targets used to determine the adjustment. We discount hazcam calibration as an error source because the hazcams were just calibrated as part of the experiment.

Regarding the random component of the error, one physical feature, seen in arc and line runs, produced the only two errors higher than 17 pixels. This feature was a small concave corner in a rock. Perhaps stereo smoothed over it, putting the recovered 3D coordinates a few centimeters too close to the navcam, accounting for 3-4 rows of additional error.

Pure kinematic handoff from far navcam

The 2D target, seen in the far navcam image when rover is 3m from the target, is handed off to the hazcams when the rover is 1m from the target. The target is projected into 3D, transformed to the hazcam coordinate system using estimates of the rover pose at 1m and 3m, and projected back into 2D using the hazcam model. Any error in the rover pose propagates into the handoff estimate.

Errors average 107 pixels, mostly consisting of projecting the target to a higher row in the hazcam image than is warranted. Each of the six runs has a relatively consistent amount of error, but there is considerable variability across runs, with means at 72, 80, 134, 178, 106, and 73 pixels. This supports the supposition that error in pose estimate – the only quantity that changes significantly across runs – is a leading cause of the handoff error. In addition, the lateral error in handoff, while consistent within each run, does not even have the same sign across runs, again suggesting a pose estimation error rather than a systematic calibration error.

The large upward offset is consistent with the rover driving farther than it expected to, bringing the target unexpectedly close and causing it to appear unexpectedly far down in the images. The rover could drive too far if it experiences less slippage than expected. Visual odometry might solve most of the pose estimation problem, but navcam and mast calibration errors would still be triple that of handoff from near navcams, as the cameras are three times farther from the target. It seems unlikely that pure kinematic handoff from far navcams would ever be superior to handoff from near navcams.

Alternate pure kinematic handoff from near navcam

In this method, JPL stereo is used to compute 3D target coordinates. This method relies more heavily on the navcam camera models because it de-warps the image and then searches along only one scan line. This could result in some error if the correct match is offset by a pixel on the neighboring line.

Average error of 22 pixels, was larger than in the pure kinematic handoff, which we would expect if error were due to miscalibration of the navcams. Again, the majority of the error consists of the predicted handoff point being above the actual handoff point in the hazcam images, consistent with miscalibrated navcams or mast.

Alternate pure kinematic handoff from far navcam

The 2D navcam target, seen from 3m away, is projected into 3D using JPL stereo, transformed to the final rover pose using the pose estimate, and projected into the hazcams.

In one run, the far navcam image was overexposed, and JPL stereo could not recover any range data. This might be considered appropriate behavior, realizing that point stereo on that same frame had produced an egregious 180 pixel handoff error. Over the remaining five runs, errors averaged 102 pixels. The errors more or less track those of point stereo, with errors in each run being comparable and errors across runs varying widely. The means were 83, 77, 146, n/a, 120, and 83. The recovered handoff point was consistently high and to the right of the point recovered using point stereo, suggesting a navcam calibration bias. This bias actually improved the handoff in the second run, where presumed pose estimate error had put the handoff point to the left of the actual target, but this improvement was purely coincidental. The average error is less than that for regular kinematic handoff, but this is just an artifact of alternate kinematic handoff's failure to track on the worst run. It seems unlikely that alternate pure kinematic handoff from far navcams would be a useful method.

Refined kinematic handoff from near navcam

The area of the left near navcam image around the 2D target is projected into 3D using JPL stereo, and then projected into the left hazcam to generate an expected hazcam target, centered on the alternate pure kinematic handoff point. This target is tracked into the actual hazcam image.

Of 18 tests (6 runs, each with three targets), 10 runs successfully tracked the target into the hazcam while the other 8 failed. In the successful runs, the average error was 1 pixel, within measurement error. Over the 8 failed runs, average error was 99 pixels.

Because the results were bimodal, it is worth examining the causes of the failures. The first two runs were fully successful. The third run, the resampled, “expected” hazcam image did not rotate the world enough to match the hazcam images. It is not immediately clear how this happened, as the transform from mast to body is reasonably well known despite small calibration errors. Nonetheless, in the third run, the expected hazcam targets did not look like the targets in the actual hazcam images, so correlation found a good but incorrect match for the first target and relatively weak matches for the others. The weakness of the matches suggests that our implementation of NCC may be buggy or at least produces counterintuitive results. In the fourth run, the targets were distorted, and tracking found the best available matches, which were not the correct matches. One of the expected targets actually was a good match for the target in the hazcam image, but the hazcam image was much darker, and an area in the lighter sand happened to be an even better match. Had the gain of the two cameras been equalized, NCC might have preferred the correct match. On the 5th and 6th runs, two of three targets tracked well. The third was a difficult target on the side of a rock, viewed at a very oblique angle. The expected target was a poor match for the actual hazcam image, perhaps because stereo is less accurate at occluding contours or perhaps because the current method for converting point clouds to expected images is weak at occluding contours. Tracking found a poor match elsewhere in the image. Had the target window been larger and lower resolution, it might have recognized the larger structure, which was well matched between expected and actual images.

The method of generating expected hazcam images from rotated navcam point clouds can produce artifacts where missing stereo data on the front of a rock leaves a hole, through which terrain behind the rock is visible. The expected images did include such holes, but none of the chosen targets fell in these areas.

Refined kinematic handoff from far navcam

In this method, the target as seen in the far navcam image is scaled down to produce an expected hazcam target, which is tracked into the actual hazcam image. This should work in the direct approach runs, where the navcams and hazcams have approximately the same line of sight to the target, and it may fail in the arced approaches where the line of sight and thus expected appearance differ.

Over 18 tests, 12 successfully tracked the target into the hazcam while 6 failed. The successes found the target with average error of 2 pixels. This is not quite as good as near refinement, which is somewhat surprising given that far handoff has fewer error sources. Perhaps the additional image detail available to the near hazcams allows construction of a better, expected target. Error failure cases averaged 83 pixels. We expected the arced runs to produce more failures than the straight line runs, because the expected and actual targets would be seen from different viewpoints. Five of the six failures did occur in arced runs, but most of these contained additional circumstances that could explain the failure, and nearly half of the arced runs succeeded.

Three of the failures were in run 4, where the far navcam image was washed out and thus did not closely resemble the hazcam image. Two features found good, coincidental matches while the third found a weak match. The method had little chance on the washed out image, but this condition occurs frequently in bright sunlight and impacts this particular method disproportionately, so it belongs in a comparative study. Gain equalization may eliminate the problem. The three remaining failures tracked to incorrect positions that did not look much like the original target. The counterintuitive matches, chosen in preference to what look like more similar and correct matches, suggest a closer examination of the correctness and suitability of the NCC tracking algorithm.

Pure geometric handoff from near navcams

In this method, hazcam pixels are projected into 3D using hazcam stereo and then reprojected into the navcam. The hazcam pixel that reprojects nearest the 2D target in the navcam becomes the handoff pixel. We expect this to be more accurate than pure kinematic handoff if hazcam stereo is more accurate than navcam stereo, which is generally the case on Rocky8.

Handoff error averaged 9 pixels. In all but three cases, geometric handoff was more accurate than pure kinematic handoff, generally by a factor of two. The exceptional cases were targets on the occluding contour of a rock. The targets were seen more obliquely in the hazcam images, so it is hardly surprising that the hazcam stereo recovery would be worse, making the geometric handoff worse. Were it not for these three cases, the range of errors

Pure geometric handoff from far navcams

This method is identical to pure geometric handoff from near navcams except that the reprojection into the navcam uses the noisy rover pose estimate. Average error was 54 pixels. This is significantly higher than handoff from near cameras, thanks to the inclusion of the noisy pose transform. However, over all test points error was less under pure geometric handoff than under pure kinematic handoff from the far cameras, thanks to superior hazcam stereo. This trend applies even to the difficult targets that foiled geometric handoff from near navcams, because the far navcams and hazcams had equally bad views of these targets and produced equally bad stereo estimates near those targets.

Refined geometric handoff from near navcams

In this method, hazcam pixels are projected into 3D using hazcam stereo, the navcam image is draped over this point cloud to provide intensity at each point, and the point cloud is reprojected into the hazcam to create an expected hazcam image. This method succeeds on 9 of 18 tests, exactly the same tests where refined kinematic handoff from near navcams succeeded, minus one. Failure on the same set of features as kinematic handoff is attributable to the fact that these features are on corners, ledges, and occluding contours – areas that are not planar. Stereo is less accurate in these non-planar areas. Further, any stereo error combines with mast/navcam pointing error to cause the navcam image to drape incorrectly. Incorrect draping on a non-planar surface causes distortion, producing expected hazcam targets that do not match the target in the actual hazcam image. The extra failure is a flat feature on the ground, but its appearance changed just enough that the tracker switched to a similar, nearby feature.

On successful tests, the method averaged 1 pixel error, within measurement accuracy. On failures, it averaged 109 pixels of error.

Refined geometric handoff from far navcams

This method is identical to refined geometric handoff from near navcams, except that the reprojection into the navcams uses the noisy rover pose estimate. This causes dramatically incorrect draping and consequent distortion of the expected hazcam targets. This method succeeded in only 3 of the 18 test cases. Those produced errors of 1, 2, and 3 pixels, for flat targets that did not distort much when draping onto other flat surfaces. For the remainder, average error was 52 pixels.

Mesh registration from near navcams

This method uses mesh registration to determine the transform that best aligns point clouds generated from navcam and hazcam stereo. This transform specifies the relationship between navcam and hazcam reference frames. The method projects the target point into 3D using navcam stereo, applies the transform to convert the target to hazcam coordinates, and projects the target into the hazcam to generate handoff coordinates. As described, mesh registration produces handoff errors on the order of 108 pixels, average. In several of the cases with the largest errors, less than 1% of the navcam mesh was retained in the alignment process, suggesting that the registration may have failed. In any case, the large error suggests failure.

To encourage the mesh registration to converge, we transformed the navcam stereo point cloud into the reference frame of the hazcams using the kinematic chain of the calibrated mast. The two meshes are thus aligned, modulo a small transform representing navcam / mast pointing error. Mesh registration was then applied to generate the small, aligning transform. Mesh registration converged in all cases, with at least 10% of the mesh retained during the alignment. The method then found the handoff point with average error of 28 pixels. The error is about double the error produced by the pure kinematic handoff that one would expect from just using the seed transform.

There may be a number of reasons why mesh registration did not produce better handoff. One particularly likely cause is a mismatch between actual and intended use of the particular implementation of mesh registration [14]. The implementation was not developed to support point clouds made using JPL camera models, and although it runs with them, it may not behave correctly. There may also be other usage nuances that were not obeyed in the test code.

A second possibility is that mesh registration cannot register meshes that have relatively small overlap, unless the meshes are first transformed into approximate alignment. This would account for the much better performance when the point clouds were initially registered using the approximate transform taken from mast calibration. It might also suggest that handoff from far navcams should be more successful, as the contents of the image are more similar to those of the hazcam images.

A third possibility is that the stereo-generated point clouds are distorted and therefore impossible to align exactly. If the cameras have gone slightly out of calibration or the rover moves between imaging, stereo would generate warped point clouds, with farther away points being more warped. Depending on how mesh registration takes this into account, it may sacrifice accuracy in the near ground to absorb significant warping in the far ground. This would account for the credible but not very precise handoff when the point clouds begin in rough alignment.

Mesh registration from far navcams

The same method is used to handoff from far navcams to hazcams, except that the navcam cloud is left in its native reference frame. The method produces handoff errors of about 124 pixels, suggesting failure to register meshes. If the navcam point cloud is first transformed to the hazcam reference frame, using the mast transform and the pose estimate, the method produces an average of 431 pixels of error, again suggesting failure. This is corroborated by the low (less than 10%) retention of points in the point clouds. The transform, being potentially quite noisy thanks to inclusion of the pose estimation, may have misaligned the meshes to such an extent that mesh registration could not compensate. The potential failures described for near navcam handoff explain most of the behavior observed for far navcam handoff. It is not clear why handoff was so much worse when the meshes were roughly aligned before handoff.

CONCLUSIONS

Principal results

The most accurate (though unreliable) handoff algorithms were refined kinematic handoff from near or far navcams and refined geometric handoff from near navcams. These methods all recover the correct handoff coordinates to within measurement error, approximately 2 pixels, in approximately half of the tests. The other half of the time, they fail completely. Refined geometric handoff from far navcams has similar accuracy when successful, but it is rarely successful. Currently, we have no means to detect success or failure, making these methods of questionable use.

Pure geometric and pure kinematic handoff, both from the near hazcams, give the next best results, averaging 9 and

13 pixel errors respectively, and succeeding in all cases. These would serve as the logical backup when refined handoff fails, and lacking a criterion to detect such failure, they are the best choice for handoff. Regrettably, they do not recover handoff within the 4-pixel or so accuracy needed to place an instrument within 1cm of the original target.

We had expected that refinement-based techniques would perform poorly on the three tests where the rover approached the target along an arc. In fact, there was no indication of this behavior.

Kinematic handoff

Near navcam, pure kinematic handoff is reasonably accurate. The error is systematic and likely due to errors in navcam pointing caused by the mast or possibly the navcams going out of calibration. Ongoing work at JPL is evaluating using a target on the rover's arm to recalibrate the mast on the fly, possibly eliminating the systematic error.

Far navcam, pure kinematic handoff is no good, probably because the estimated rover motion is inaccurate. Visual odometry might reduce this error, but error is always likely to be higher from navcam stereo taken at a greater distance, so there is no obvious reason not to prefer near-navcam pure kinematic handoff.

Alternate, pure kinematic handoff, which uses JPL stereo instead of point stereo, is less accurate in all cases except one case where the extra error cancelled error from another sourc. Most likely the alternate handoff was worse because JPL stereo is heavily optimized to rely on correct camera models, and the navcam models are probably out-of-date.

Refined kinematic handoff works extremely well half the time and extremely poorly the other half. Handoff from near navcams is slightly better, but both are within measurement error. The same algorithm had been less accurate (3 to 4 pixel errors) but more reliable (75% success) in earlier in the program, suggesting perhaps that the targets for the current tests were more polarized into easy and difficult.

Failures in refined kinematic handoff seem to be caused by poor projection of near navcam images or washed out far navcam images. Refinement might work more reliably on larger, lower resolution features that include more content. This could be followed by a second search over a small area at full resolution, to complete the refinement. This was the original strategy used in the SCIP implementation. Refinement might also work better if the gain around the targets were equalized, giving the expected and actual targets comparable intensity. That might make the actual target more attractive than a feature in the sand that has less similarity in contrast but more in intensity. Also, based on some of the suspicious matches made by the normalized cross correlator, it would be a good idea to verify the correctness of the NCC code, the suitability of NCC for responding to lighting changes, and whether the expected target window is being built correctly.

Geometric handoff

Pure geometric handoff from near hazcams is better than pure kinematic handoff from the same cameras whenever hazcam stereo is better than navcam stereo. This is always the case except when navcams view an upward-facing feature from above while the hazcams view it obliquely. The error sources are the same in both cases, but the hazcam stereo is likely more accurate because the hazcam calibration is more recent.

For the near navcams, refined geometric handoff is comparable to refined kinematic handoff in both accuracy and computation. However, note that if pure geometric handoff is to be done as a backup, then most of the computation for refined geometric handoff is already done.

For the far navcams, any geometric handoff is useless, as the noisy pose transform causes the navcam image to drape incorrectly over the hazcam stereo, producing incorrect expected hazcam features. In incorrect draping is clearly visible in the expected images. Perhaps visual odometry could remedy the pose error, the resulting draping error, and resulting error in expected target, and the resulting error in handoff. In three of eighteen cases, geometric handoff from far navcams draped the target over compatible (though incorrect) parts of the hazcam stereo and generated visually correct expected targets that tracked correctly. That is hardly the basis for a reliable system. And, as with refined kinematic handoff, correlation peak height does not correlate with successful tracking.

Mesh Registration

The mesh registration implementation used in the tests did not perform to the level anticipated based on results reported in the literature. Handoff results were credible only when meshes were roughly aligned before applying mesh registration. In those cases, mesh registration made the alignment worse. This is consistent with [10], which found that ICP performed mesh registration well on simulated data but not on point clouds generated using JPL stereo.

Limitations

This work was intended to be a quick implementation / investigation of the refined geometic handoff technique described in [13]. It has grown into a more systematic study and is presented in hopes that it will be useful as a benchmark against which to evaluate additional handoff algorithms. Given the limited, intended scope, there are a number of avenues that were not explored.

The work does not attempt to exhaustively sample the field of possible handoff algorithms. Most notably, it does not consider algorithms by which the rover actively places fiducials in a scene and then images and evaluates them to determine navcam-to-hazcam coordinate transform.

The work does not compare the actual speed of algorithms, as none of the implementations tested are currently optimized for speed. Instead, the discussion of handoff methods gives some indication of the relative speeds by comparing the amount of computation required.

Ground truth for handoff positions is measured by visual inspection of the hazcam images, so the reported average errors may be off by one or two pixels. The NCC algorithm used with the refinement methods currently does not track to subpixel accuracy, so results may be off by an additional pixel. This is still sufficient to identify the trends in the algorithm. For comparison, a 1cm error on a target 1m from the camera, often considered acceptable error for SCIP, corresponds to about a 6 pixel error in the hazcams, well above the precision of the experiments.

The cameras on Rocky8 are not synchronized. Should the cameras move between image capture by the various cameras, then the camera models will be in different reference frames. This will introduce error in stereo processing and conversion between navcam and hazcam reference frames. Such movement has been observed in pancam images, as the rover brakes, the suspension rocks while damping out the deceleration, and the rocking is amplified along the camera mast. We now wait approximately 20 seconds between braking and imaging, to let any motion damp out. Nonetheless, if the mast moves, point clouds generated from stereo and transforms based on mast calibration may contain error that reduce the reported handoff accuracy.

The pose estimation used during SCIP is not particularly accurate, being based principally on wheel odometry. This may unfairly penalize the far-navcam handoff methods

Future work

Given the preceding results and observations on probable error sources, we can posit the following directions that might lead to improvements in the handoff methods described here.

Camera synchronization—Synchronizing cameras prevents the rover from moving between taking images and might improve stereo and/or mast calibration (the estimated transform from navcam to hazcam coordinates). More likely, any such motion is already damped out during an intentional delay before taking images. However, synchronization removes this source of error from consideration, and it allows handoff to proceed without the inserted delay.

Mast Calibration—Ongoing work at JPL is developing a mechanism to adjust mast calibration on the fly using targets on the rover's manipulator arm. Preliminary results show reducing pure geometric handoff error from 11 to 3 pixels. Presumably pure kinematic handoff would receive similar benefit.

Visual odometry—This could aid the various far-navcam handoff methods in two ways. First, it could provide an accurate estimate of rover motion, eliminating noise in the conversion from navcam to hazcam coordinates. Second, it might be used to recover the transform between hazcam and navcam frames, as mesh registration does, but using sparse features. This should require a translation/scale tracker, as the two image sets would have a similar line of sight to the targets, but have different distances and different focal lengths. Finding the transform using visual odometry should be faster than with mesh registration because it would use only sparse features. It could also be more accurate, as the features could be chosen near the pure kinematic handoff point, so that even were the stereo skewed by poor camera models, the transform would properly represent the relevant part of the images.

Wider Arc—We tested handoff from far navcams to hazcams when the two sets of cameras shared a similar line of sight, and when their lines of sight were offset by about 30 degrees. The idea was to show that far kinematic handoff would fail on an angled approach, where the change in the appearance of the 2D target could not be explained by a single scaling factor. If there was any such trend, it was hidden by larger error sources. Consider another experiment where the rover moves in a circle around the target, taking images from different angles. Evaluate the maximum angle for which the various methods can still recover the handoff point.

Failure detection—Identify whether refinement has succeeded. Refined kinematic and geometric handoff were the only methods with the handoff accuracy required for SCIP, but they are not reliable. In the 18 test cases in this work, and in previous work, there has been no correlation between success of the NCC tracker and the NCC correlation peak. This is reasonable, because in many cases, an incorrect match is visually more similar to the original target. This might suggest that NCC is not the appropriate tracker, or that it is buggy, or that we must track larger features that have more context, perhaps at lower resolution. Perhaps after fixing some of these items, it will be possible to threshold a correlation peak to identify success. An unappealing alternative is to threshold based on the distance between the handoff coordinates determined by refinement and those determined by pure kinematic or geometric handoff. This is unappealing because, if there existed such a threshold distance, it should be used to limit the search area and thus time required, not test for success. Perhaps the smaller area would include fewer false matches, so that correlation peak would be a better indicator of success.

Subpixel accuracy—If refinement becomes reliable and thus usable, it may be useful to handoff to subpixel accuracy. The tracker currently has its subpixel interpolation disabled, because it is not robust to the gain change between the navcam and hazcam images. Were this problem solved, the handoff might be found to be even more accurate, perhaps allowing one pixel across the entire SCIP run rather than just across handoff. Such tests would require better ground truth measurements, and they would also warrant some preliminary testing to verify that the other elements of SCIP have similarly low contributions to total error.

Mesh registration—Investigate why mesh registration was unable to recover large transforms between meshes and unable to improve rough alignment. Probable reasons include incompatibility with point clouds from JPL cameras, inability to weight meshes to accommodate point clouds produced from miscalibrated cameras, and inability to register meshes with significant non-overlap when the meshes do not begin in rough alignment.

Reverse handoff—Generate hazcam stereo, with each point colored according to the hazcam pixel(s) that it represents. Project that point cloud into the navcam to make an expected image. Track the navcam 2D feature from the actual navcam image into the expected image. Kinematically project the resulting point back into the hazcam. This is essentially the reverse of refined kinematic handoff. It should be better if hazcam stereo is better than navcam stereo. It should be better than geometric handoff because it does not rely on perfect mast calibration to color the hazcam stereo. This method would probably only work for the far navcams, which have comparable resolution to the hazcams and so could create a dense, expected navcam image. As with refined kinematic handoff, the expected image may have artifacts that could cause tracking failure.

ACKNOWLEDGEMENT

The work described in this paper was carried out at the Jet Propulsion Laboratory, California Institute of Technology and funded by the Mars Technology Program.

REFERENCES

[1] Liam Pederson et al, "Mission Planning and Target Tracking for Autonomous Instrument Placement", IEEE Aerospace Conference, Big Sky, Montana, March 2005.

[2] Paul Backes, et al, "Automated Rover Positioning and Instrument Placement," IEEE Aerospace Conference, Big Sky, Montana, March 2005.

[3] Terry Huntsberger, et al, "Closed Loop Control for Autonomous Approach and Placement of Science Instruments by Planetary Rovers," *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2005),* Edmondton, Alberta, Canada, August, 2005.

[4] Max Bajracharya, et al, "Target Tracking, Approach, and Camera Handoff for Automated Instrument Placement," IEEE Aerospace Conference, Big Sky, Montana, March 2005.

[5] Liam Pedersen, et al, "Multiple-Target Single Cycle Instrument Placement," iSAIRAS 2005.

[6] Larry Matthies, Dynamic Stereo Vision, Carnegie Mellon thesis CMU-CS-89-195, October 1989.

[7] Matt Deans, et al, "Combined feature based and shape based visual target tracker for robot navigation", IEEE Aerospace Conference, Big Sky, Montana, March 2005.

[8] Matt Deans, et al, "Terrain model registration for single cycle instrument placement", in IROS 2003.

[9] Paul Besl and Neil McKay, "A Method for Registration of 3-D Shapes," PAMI 14:2 February 1992.

[10] Richard Madison, "Analysis of ICP for 3D Feature", http://keuka.jpl.nasa.gov/main/software/technology/visual_tracking/ICP_report_madison.doc, November 2003.

[11] Issa Nesnas, et al, "Visual Target Tracking for Rover-based Planetary Exploration," IEEE Aerospace Conference, Big Sky, Montana, March 6-14, 2004.

[12] Rocky 8 calibration website http://keuka.jpl.nasa.gov/main/testbed/how_to/calibrate.html.

[13] Won Soo Kim, Adnan Ansar, and Rob Steele, "Rover Mast Calibration, Exact Camera Pointing, and Camera Handoff for Visual Target Tracking," IEEE International Conference on Advanced Robotics (ICAR), May 2005.

[14] Mesh registration 2005 delivery website http://keuka.jpl.nasa.gov/main/software/technology/mesh_registration_arc/index.html.

BIOGRAPHY

Dr. Richard Madison *is a Member of the Technical Staff in the Computer Vision Group at the Jet Propulsion Laboratory. He earned his B.S. degree in Engineering at the Harvey Mudd College in 1992, and his M.S. and PhD degrees in Electrical and Computer Engineering in 1994 and 1997 from Carnegie-Mellon University. He works on computer vision related infrastructure tasks for the Mobility and Robotics Systems Section at JPL.*