Space 2006
19 - 21 September 2006, San Jose, California

AIAA 2006-7385

Performance Evaluation of Handoff for Instrument Placement

Liam Pedersen*, Clay Kunz, Randy Sargent
Carnegie Mellon University / NASA ARC, Moffett Field, CA 94035, USA

Richard Madison†, Paul Backes, Max Bajracharya, Daniel S. Clouse, Issa Nesnas
Jet Propulsion Laboratory, Pasadena, CA 91109, USA

and

Matthew Deans
USRA / NASA ARC, Moffett Field, CA 94035, USA

[Abstract] Single Cycle Instrument Placement (SCIP), the ability to autonomously approach rocks and place contact instruments within 1cm of selected features, has been identified as a high priority for increasing the efficiency and science return for planetary rover surface operations. Because of imprecise localization, it is necessary for a rover to visually keep track of targets, using images from onboard stereo cameras, as the rover navigates to them. "Handoff" is the problem of matching a tracked point in one camera to the corresponding point in another camera pair. Handoff facilitates visual tracking of a point using multiple camera pairs, allowing at any time the choice of camera pair with the best view of the portion of the scene that is of interest. A significant fraction of the tracking error budget is due to handoff. In this paper, we review several methods for solving what we refer to as the handoff problem and evaluate their performance in the context of integrated planetary rover test beds analogous to the MER Spirit and Opportunity vehicles.

I. Introduction

The ability to autonomously approach rocks and place contact instruments on selected features has been identified as a high priority for increasing the efficiency and science return for planetary rover surface operations. Robotic activity on Mars is not amenable to tele-operation 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

* Systems Scientist, Intelligent Robotics Group, NASA Ames Research Center, MS 269-3, Moffett Field, CA 94035-1000, Member AIAA.
†IS/CS, Jet Propulsion Laboratory, MS 125-209, 4800 Oak Grove Dr, Pasadena, CA 91109-8099

Copyright © 2006 by the American Institute of Aeronautics and Astronautics, Inc.
The U.S. Government has a royalty-free license to exercise all rights under the copyright claimed herein for Governmental purposes.
All other rights are reserved by the copyright owner.

operations as well as similar pick-and-place operations such as sample acquisition, resource collection, return to a processing facility, and construction.

Recent research efforts at both NASA's Ames Research Center and Jet Propulsion Laboratory have made significant advances in these capabilities within the context of integrated demonstrations of remote science operations, visualization and situational awareness, planning and execution, and robot vision, navigation, and control.

In this paper, we review several methods for solving what we refer to as the "handoff" problem and evaluate their performance. Rovers such as the MER Spirit and Opportunity, JPL's Fido and Rocky8, and Ames's K9 rover have several calibrated stereo camera pairs onboard with differences in mounting location, field of view, and resolution. Handoff is the problem of matching a tracked point in one camera pair to the corresponding point in another camera pair. Handoff facilitates visual tracking of a point using multiple camera pairs, allowing at any time the choice of camera pair with the best view of the portion of the scene that is of interest.

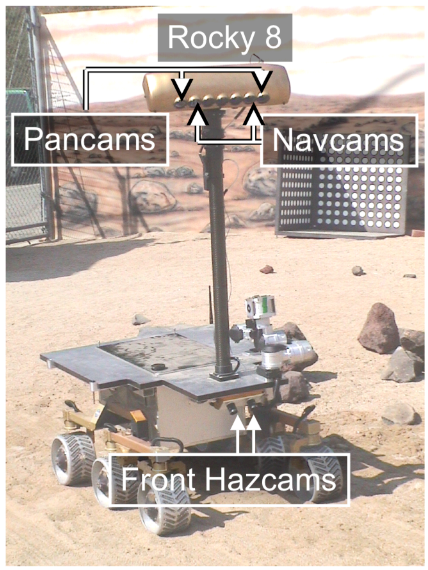


Figure 1 : *JPL's Rocky 8 planetary rover testbed, indicating the cameras used for initial target designation (pancams, 6° fov), target tracking (pancams, navcams, hazcams) during approach, final instrument placement (hazcams) and obstacle avoidance (hazcams).*

In a typical target approach and instrument placement run, the placement point is initially designated from panoramas taken with the high resolution mast cameras (SciCams or PanCams; Figure 1), which provide the best view of points several meters distant. At the final rover position, when the rover is ready to use the manipulator to place the instrument, the imagers mounted low on the front of the rover (HazCams) typically have the best view of the intended instrument placement point and arm workspace. During rover approach navigation there may also be advantages to using the midrange navigation cameras (NavCams) to track the target. Thus, in a typical approach and instrument placement, camera handoff will be required at least once, and often twice. In multiple target single command activity, handoff could happen several times. Each handoff is a source of error, and these errors are compounded along with visual tracking errors to limit the overall accuracy of the system. Understanding the sources of errors and reducing errors as much as possible is an important step toward maximizing the utility of autonomous approach and placement.

The performance evaluation in this paper considers several characteristics of the algorithms reviewed, including discussions of the computation required as well as known limitations of various approaches with respect to real world issues such as robustness to changes in illumination or viewpoint. Performance is measured in terms of image plane distance measured in pixels. Success and failure rates for methods are also compared. Tests are performed using imagery from the Mars Yard test facility at JPL.

II. SCIP Implementations

Variations of SCIP have been and continue to be developed and demonstrated, starting with [1] at NASA ARC and continuing with [2][3][4] at JPL, using differing approaches to the target approach and tracking problem.

A. JPL's Rocky 8 Rover SCIP Implementation

The ongoing work at SCIP work at JPL [4] is improving components from several of earlier efforts and 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 the Mars Science Laboratory and other, future Mars missions.

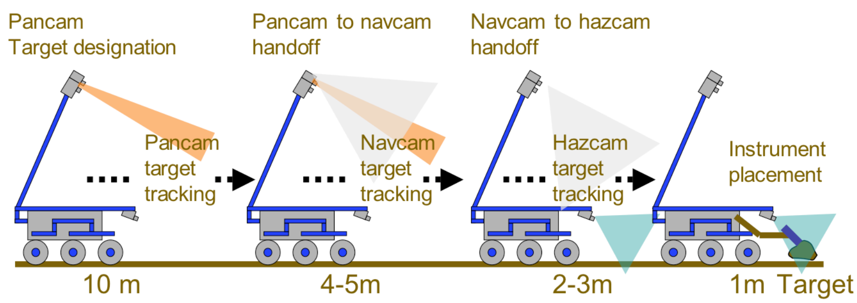


Figure 2 *JPL SCIP approach sequence*

The high-level algorithm of this version of SCIP, running on the Rocky8 rover, consists of the following steps.

(1) User designates a 2D target in a high resolution pancam image from 10m. Pancams see ~3mm/pixel at 10m, allowing precise target designation. Navcam resolution at 10m is ~1cm is insufficient for target designation. Stereo processing identifies the target's 3D location. A 2D template of the target area is created for the rover to track the target.

(2) Rover drives midway to target, stopping frequently to check for obstacles to get new pancam images with which to track the target [11]. Tracking is done by matching an affine transform of the 2D template image of the target, using the normalized cross correlation (NCC) similarity metric, and then using stereo to update the target 3D position. A requirement of this tracker is that the target viewpoint not change considerably from the template image.

Note that targets location with respect to robot becomes more accurately determined the closer it gets (assuming that tracking errors do not accumulate).

(3) At 4-5m from target, pancams have too narrow field of view (6°) to be reliably pointed at target, which is therefore handed off to the navcam cameras which have 3x the field of view, and therefore 3x the tolerance to pointing error. (At further ranges, the reduced resolution of the navcams would lead to greater tracking error). Rover continues to target, using navcam images to track target location.

(4) At 2-3m from the target, the navcam view of target becomes too steep for the tracker to work well and

the navcams cannot be easily pointed at it. The target is therefore handed off to the front mounted hazcams, which now have a more similar view to the original view of the target, and are now used to track the target.

(5) At 1m from the target, hazcam images are used to get a final 3D position fix for it, and the rover's manipulator arm places an instrument against it.

On the Rocky8 rover, and by extension on the similarly-configured MER rovers, handoff from pancams to navcams is straightforward. A window around the 2D target in the left pancam image is shrunk 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. This view is used as the tracking template for subsequent navcam tracking updates. 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. This 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.

The navcam to hazcam handoff is considerably more difficult as the two camera pairs do not share a similar viewpoint. The rover creates a stereo 3D point cloud from the navcam images and projects this into the hazcam frame to create an "expected" hazcam 2D 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 this handoff error.

B. NASA ARC K9 Robot SCIP Architecture

The Ames SCIP implementation [1][5] on the K9 rover differs in the target tracking algorithm during the approach phase, and uses only a single hand-off (to hazcams) when the rover is in front of the target rock:

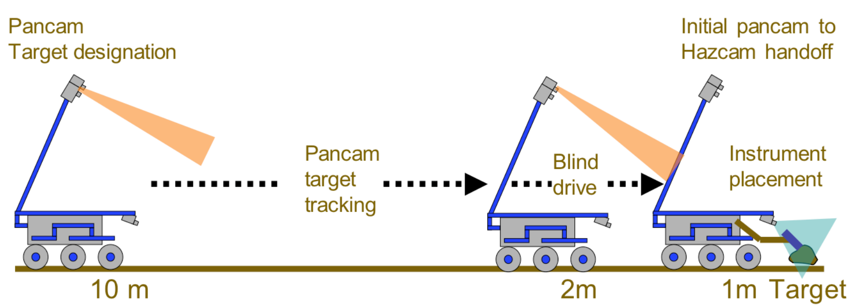


Figure 3 *ARC SCIP approach sequence*

(1) The rover (K9) generates a stereo image panorama using the high resolution (6° fov), mast mounted pancams (called scicams on K9). These are processed to create a 3D model with the ARC Stereo Pipeline [Ref], within which a user designates 3D target points and cropped 3D template models of the target areas. The cropped 3D template is typically 25% x 25% of the 3D model from a single stereo pair.

(2) Every 0.5m the rover stops, to detect obstacles and get pancam (scicam) stereo image pairs of each of the remaining target points, which it uses to update the apparent target 3D locations. The rover does this for as long as the target is in view (about 2m) and then dead reckons the final 1m to the target. The ARC target tracker is very robust to extreme changes in viewpoint, partial occlusions and requires only limited image overlap with the previous set of tracking images in order to update the target position. Hence no midcourse pancam to navcam handoff is necessary (or desirable).

(3) Once within reach of the 1st target, the rover acquires hazcam stereo images of the target,

computes a 3D model of the target area from them, and registers [7] the 3D template acquired in (1) to this, thus computing the 6DOF transform from the point specified in the ***original*** images to the current hazcam model.

(4) After an onboard safety check [1][5], the rover uses its manipulator arm to place the instrument on the target.

(5) The rover stops tracking the current target, and moves on to subsequent targets.

This system was designed to allow SCIP on *multiple* targets. Key technical distinctions from the Rocky 8 system are the visual tracker and the hand-off algorithm. The tracker [REF] works by matching 3D interest points ("SIFT keypoints") between image pairs, and computing the apparent 6DOF motion of the target between rover moves (this apparent motion is due to inaccuracies in the rover and initial target localization). The apparent location (in the local site coordinate system) of the targets is updated at each step with the addition of its apparent motion. This tracker accuracy is about 1.3 – 1.7cm after a 10m approach [5]), but the error increases monotonically over traverse distance. Since the tracker uses large numbers of interest points dispersed around the target area, it is robust to occlusions of the target itself. Besides allowing the rover to circumnavigate targets (a possibility in the muli-target SCIP scenario), this means that tracking images need not contain the target, instead they must simply overlap with at least 50% of the previous tracking images. This diminishes the pointing accuracy requirements and allows the ARC tracking system to use the high resolution pancams (science cams) exclusively, which is fortunate because they are the only cameras able to resolve sufficient texture around the target area to recover robust interest points there.

The main drawback (besides computational burden) of the tracking system is the unbounded growth in the tracking error after each update. The mesh registration handoff system recovers from this by matching templates from the original images to the current hazcam images. Therefore, provided hand-off succeeds, the entire tracking and approach error budget is given by handoff error only.

III. Handoff Methods

Our experience is that target hand-off error dominates the overall target approach phase error. This is true for both JPL and ARC SCIP architectures described in section II. The ultimate goal is 1cm ***total*** error. Furthermore, on MER, 0.5cm of the total instrument placement error budget is allocated to manipulator errors [15], although the ARC system comes closest with approximate approach phase accuracies of 1.6 to 2.7cm [5]. This motivates our analysis of a broad range of hand-off methods, which we describe and evaluate in this paper.

A. Pure kinematic handoff

This uses 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 consists of the *mast calibration*, the transform from navcams to hazcams at one rover pose, possibly combined with the pose change of the rover after taking the navcam images.

The target is projected into 3D coordinates using point stereo, which takes a target in the left image and finds a match in the right 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 calibration (poor calibration or vibrating mast), incorrect hazcam model, and when moving rover between navcam and hazcam images, incorrect relative pose estimation. Mast vibration is an a big issue if the cameras are not synchronized, as any motion between getting left and right images invalidates their calibration.

B. Alternate pure kinematic handoff

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.

C. Refined kinematic handoff

The next two methods expand on kinematic handoff by projecting not just the navcam target but a window around it into 3D and then through the hazcam model to produce an expected hazcam target image. They track this expected target into the actual hazcam image using normalized cross correlation (NCC) [11]. Normalization is important, as the autogain of the cameras can cause significant intensity differences between the navcam and hazcam images.

The expected hazcam target image is determined as follows:

(1) Use JPL stereo on navcam images to build a point cloud. Each point takes the intensity of the associated left-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, consider the point when filling each of the four pixels. With this algorithm, the expected image may contain holes through which points from the far side of the point cloud may be visible, possibly causing errors in the tracking.

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

If rover motion occurred between hazcam and navcam images, the expected hazcam image is scaled by the ratios of fields-of-view and distances-to-target for the two camera sets. This motion may help give both cameras the same perspective on the target, improving tracker reliability. 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, this requires 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 slows 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 close to the target, 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 a low correlation peak does not always betray the failure. 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.

D. Pure geometric handoff

This method generates a point cloud from hazcam stereo, projects it into the navcam images, and identifies the point that projects nearest the navcam 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, they sit at the top of a long and possibly moving lever arm, and they are farther from the target. Geometric handoff can be done from near or far navcams, but projecting into far navcams requires use of the noisy, rover pose estimate.

E. Refined geometric handoff

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. Hazcam pixels are projected into a 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 intensities at those coordinates are copied back to the hazcam pixels. 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 to drape incorrectly over the hazcam stereo, perhaps 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 rover pose estimate is very accurate or the target is on a large, flat area.

As with refined kinematic handoff, it might be preferable to generate the expected feature using a rectified hazcam model, rectify the actual hazcam image, track, and then unrectify the resulting handoff point. In practice, this is unnecessary if the target is near the center of the image.

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.

F. Mesh registration

Mesh registration [7][8] works by aligning 3D point clouds obtained from the two camera pairs (pancams/navcams and hazcams) to recover the transform between them. The 3D position of target points in one set of image is computed using JPL stereo, mapped using this transform to the hazard camera frame, then projected to hazard camera image plane to recover the 2D coordinates.

Mesh registration is an improvement on the Iterative Closest Point (ICP) [9] algorithm for determining the transform that best aligns two point clouds. Firstly, it assumes good orientation information, and does an initial correlation search over a range of translations orthogonal to the hazcam optical axis. This is used to seed a Nelder-Mead optimization of the full 6 DOF transform between the two models. Secondly, mesh registration establishes point correspondences between models based on their projection onto the hazcam image plane (Figure 4). The correspondance can be established in constant time, unlike ICP, so that overall mesh registration CPU time is comparable to a stereo computation, but slower compared to the other methods here.

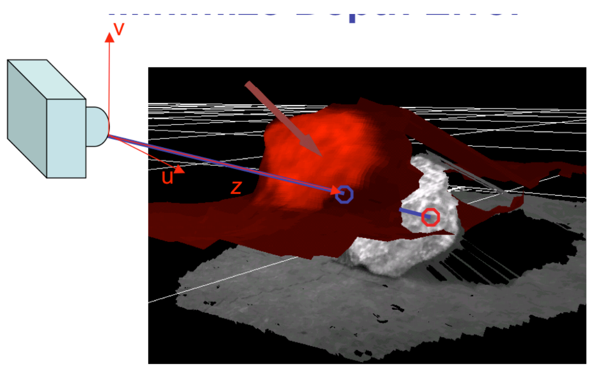


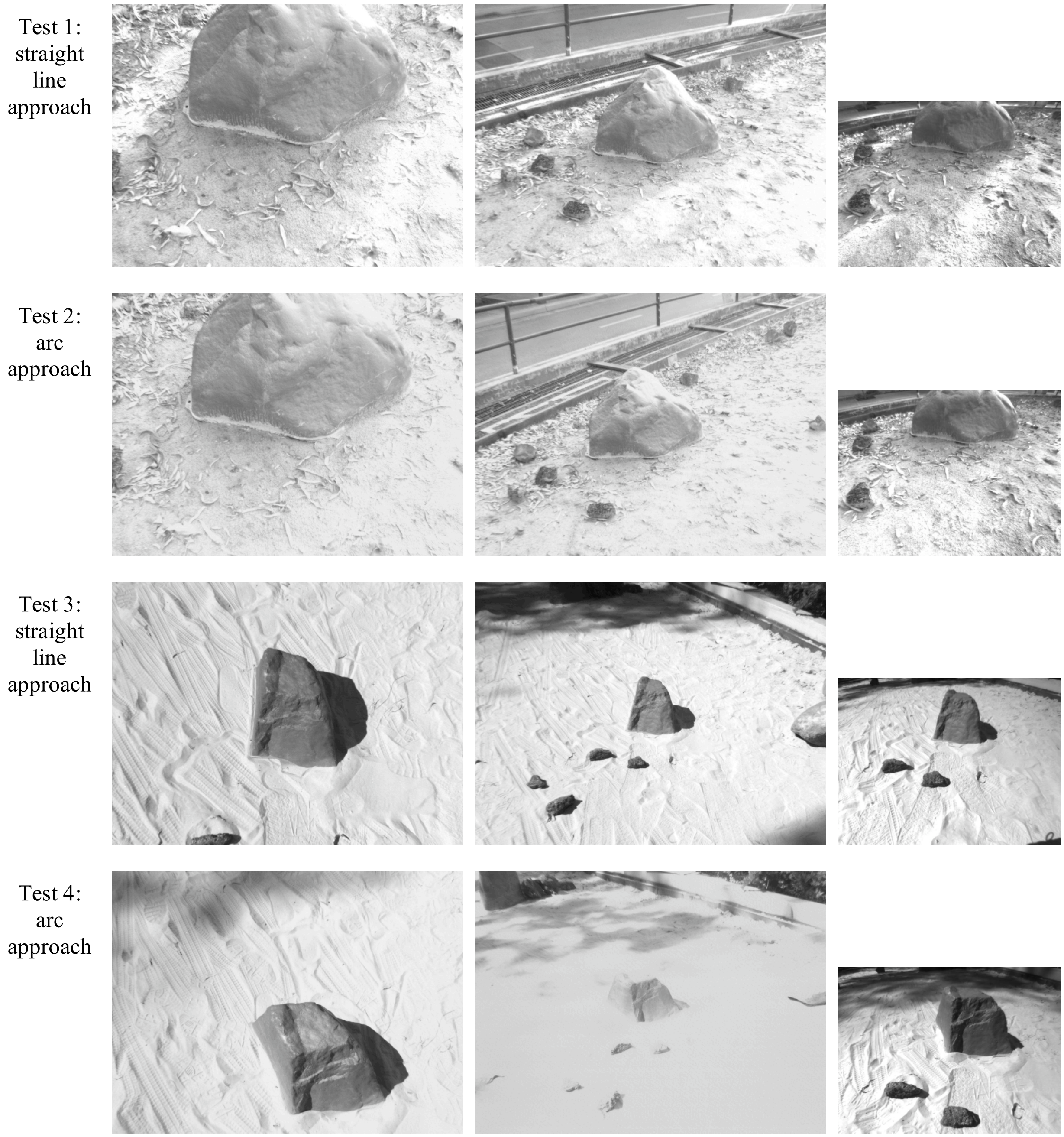
Figure 4 *Mesh registration algorithm. Cropped pancam (or navcam) 3D stereo models are transformed into the hazcam coordinate frame using mast calibration and rover motion estimates. Point correspondences are established by projecting from the camera origin. An initial search along (u,v) finds the approximate lateral translation maximizing the normalized cross correlation of the camera - model ranges. Nelder-Mead minimization to find the full 6DOF transform that minimizes the sum of square distances between corresponding points follows.*

Mesh registration relies on a reasonable initial alignment of the models, based either on rover localization, target tracking information or both. In these tests, we seed it with initial alignments based on inaccurate rover pose estimates only.

Imperfect stereo camera calibration leads to 3D models from stereo images that are only locally accurate. As this is invariably the case, registering full sized 3D stereo models is unlikely to yield the optimal 6DOF transform for target points. For this reason, and to increase speed, handoff with mesh registration is done using navcam (or pancam) stereo models cropped around the target points. This is done manually for pancam images (within which users select the target points in the first place). For handoff from close up navcam to hazcam images (as required by the JPL architecture), the cropping parameters need to be inferred from the pancam cropping parameters. Experience indicates that it is not especially sensitive to these. Our best practice is to ensure that the target rock is contained within the cropped region, along with the ground about 1 rock radii on either side of the rock.

IV. Experiment

To evaluate the relative merit of the handoff methods, we performed each of the types of handoff on target points, chosen within the *near* (1m ground distance) navcam images, and *far* (3m ground distance) navcam images to hazcam images. The targets 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 Figure 2. All images were taken with the Rocky 8 rover, at JPL's Marsyard. The experimental procedure was to place the rover in front of each of the 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 navcams and hazcams. Then 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 navcam images. At this distance, the navcams have approximately the same line of sight to the large rock as the hazcams had when we took hazcam images (Figure 3). 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. The image sets simulate pairs of approaches to the same target, along straight-line and arcing paths.


Test 1: straight line approach
Test 2: arc approach
Test 3: straight line approach
Test 4: arc approach

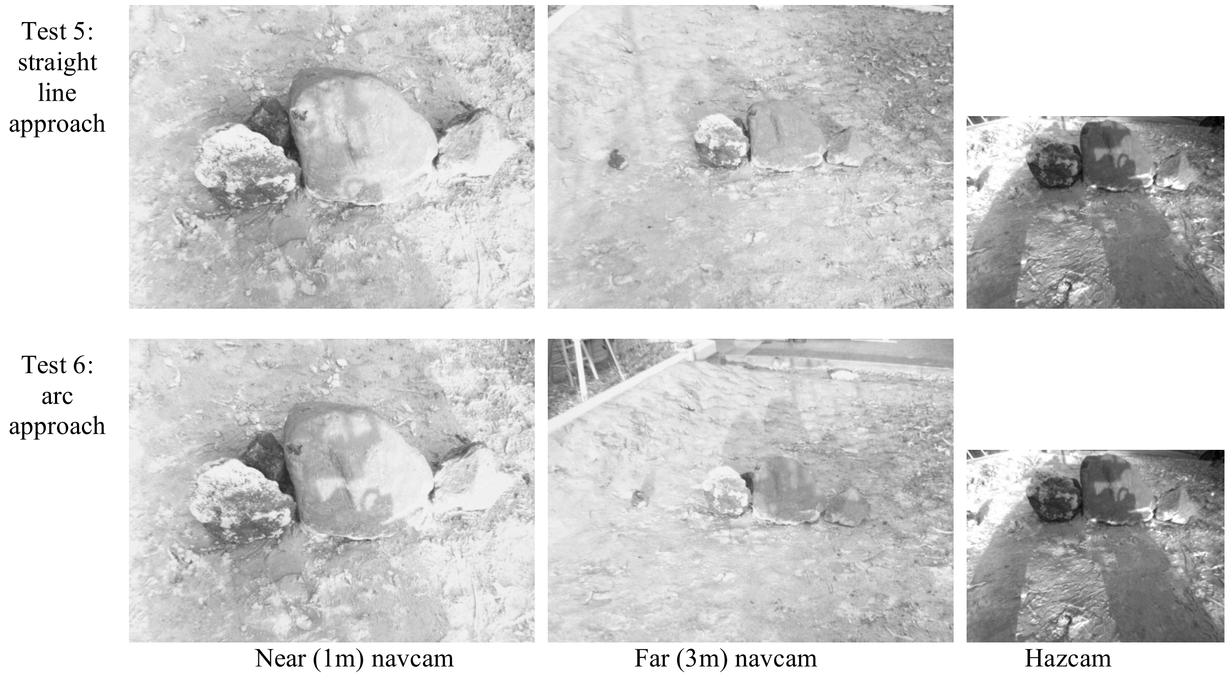


Figure 5 *The left near navcam (a), far navcam (b), and hazcam (c) images used in test runs. Note that Far navcam images have approximately the same perspective as the hazcam images (Figure 3), although far navcam images taken "prior" to arc approaches differ by a 30° yaw rotation. The far navcam image in test 4 is badly washed out, but handoff algorithms must accommodate this common situation.*

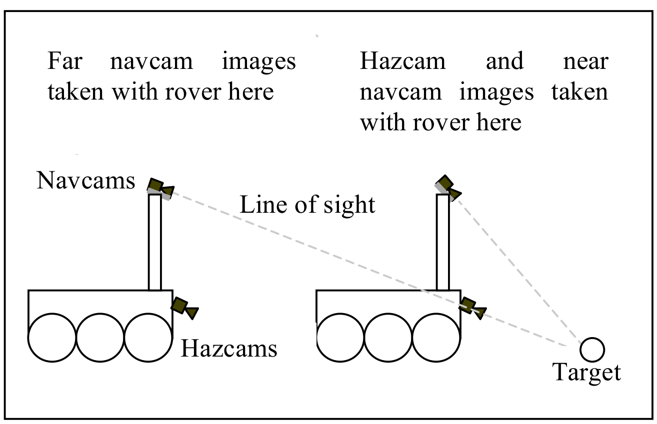


Figure 3 – Hazcam, near and far navcam image locations. Note similar perspective (differing mainly in scale) of far navcams and hazcams

Including the far navcam images allows us to control for the effects of a change in viewpoint on the hand-off algorithms performance. The rover pose transformation between the far and near positions is based on integrating commanded wheel moves (to get x,y and yaw changes), and a noisy IMU (for roll and pitch).

For each of the six tests, we manually identified corresponding target points that appeared in near navcam, far navcam, and hazcam images (Figure 4). The goal was to have targets on flat places, sharp points, occluding contours, sand, and high contrast points. A shortcoming with this technique is that all chosen points need some amount of texture to be matched, possibly introducing a bias in the handoff results for algorithms that use texture.

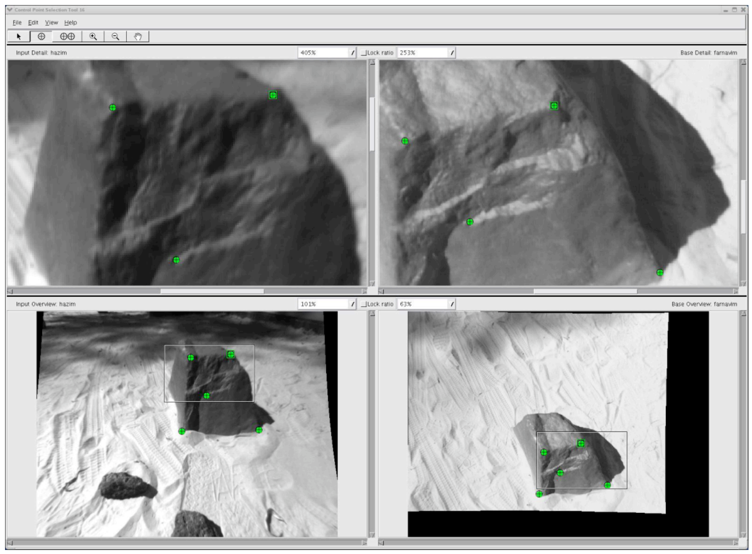


Figure 4 – Ground truth corresponding target points in far navcam and hazcam images, picked manually using natural texture.

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 navcam reference frame. 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.

A note on the mesh registration experiments. A single cropped template was created for each navcam image, enclosing the target rock and approximately 1 rock radii on either side. This was registered with the hazcam 3D model, and the computed transform used to map the navcam target points to their corresponding locations in the hazcam images. The process was repeated using uncropped navcam images as template models for comparison purposes.

V. Results and Discussion

Table 1. Summary of handoff accuracies and success rates *under Rocky 8 operating circumstances, including limited localization accuracy but excluding failures in stereo correlation.* 1 pixel error ≈ 0.25 cm

Method	Navcam	Success	Average error (pixels) ± standard deviation	Error range (pixels)
Pure kinematic	Near	100%	16 ± 4	3-22
	Far	100%	99 ± 26	70-156
Alternate pure kinematic	Near	96%	13 ± 2 / 59 ± 84	10-16 / 99-218
	Far	100%	96 ± 26	67-154
Refined kinematic (success / failure)	Near	79%	3 ± 2 / 39 ± 25	1-9 / 16-65
	Far	68%	1 ± 1 / 37 ± 21	0-3 / 17-70
Pure geometric	Near	100%	8 ± 5	2-21
	Far	100%	59 ± 24	16-120
Refined geometric (success / failure)	Near	84%	1 ± 1 / 30 ± 19	0-4 / 8-45
	Far	12%	66 ± 45	1-170
Mesh registration	Near	100%	4.3±2.2	0.75-8.5
	Far	60%	4.9±5.5	0.75-19
Mesh registration using uncropped navcam models	Near	100%	14 ± 11	2.2-55
	Far	100%	49 ± 30	8.2-108

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, *in the context of the overall rover system with the limited localization accuracy and failures in stereo correlation.* 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 reliable. Among the reliable methods, mesh registration with cropped templates gives the best results, averaging 4.3 and 4.9 pixel errors for the near and far navcam to hazcam hand-offs respectively.

Pure kinematic handoff from near navcams

Errors averaged 16 pixels, with the actual target in the hazcam images being consistently about 15 rows below the predicted position. Random error accounted for up to 6 pixels of error around this mean. A single outlier inexplicably had only 3 pixels of error total.

The systematic nature of the error 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%. We have previously observed cameras going out of calibration enough to cause a 5% ranging error.

Another explanation is incorrect mast calibration. Unpublished work at JPL on autonomous mast calibration adjustment found a nearly 1 degree tilt error in mast calibration, which is enough to produce the observed 15 pixel offset. There is no obvious mechanism by which the mast would go out of calibration by a full degree, but even if mast calibration adjustment corrects for other errors, at least in the neighborhood of the target, it may be acceptable for supporting handoff.

We discount hazcam calibration as an error source because the hazcams were just calibrated as part of the experiment.

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. More regular calibration of the mast might eliminate this error.

A. Pure kinematic handoff from far navcams

The 2D target is identified in the far navcam image when rover is 3m from the 3D target and then handed off to the hazcams when the rover is 1m from the 3D target. The 2D navcam target is projected into 3D when it is first located, transformed to the final 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 99 pixels, mostly consisting of projecting the target about 100 rows higher in the hazcam image than is warranted. The errors for features within a given run are consistent, with standard deviations around 5 pixels. However, there is considerable variability in mean error across runs, with means from about 65 to about 145 pixels.. This supports the supposition that error in pose estimate – the only quantity that changes significantly across runs – is a leading source 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. Even so, errors from navcam and mast calibration would be triple those of handoff from near navcams, as the far navcams 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.

B. Alternate pure kinematic handoff from near navcams

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, or it could be more accurate if camera models are accurate.

The method succeeded in all but two cases. When successful, average error was 13 pixels, slightly better than in the pure kinematic handoff, but probably within measurement error. 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. For two features, handoff errors were 99 and 218 pixels, suggesting complete failure. The features show no obvious reason why stereo should fail, but one of the features is the feature for which pure kinematic handoff from near hazcams produced its single outlier error result.

C. Alternate pure kinematic handoff from far navcams

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. Behavior was nearly identical to that of pure kinematic handoff from far navcams. All runs succeed, though mean errors are round 96 pixels. Each test has a consistent mean error and small deviation for the features of that run. The large mean errors are essentially the same as pure kinematic handoff. It seems unlikely that alternate pure kinematic handoff from far navcams would be a useful method.

D. Refined kinematic handoff from near navcams

The area around the 2D target in the left near navcam image 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 25 tests (5 runs, 5 targets each), 20 tests successfully tracked the target into the hazcam, 4 lost tracking completely, and 1 target was too close to the image edge to produce a window suitable for tracking. In the successful runs, the average error was 3 pixels, almost within measurement error. Over the failed runs, the average error was 38 pixels. The accuracy of the successful handoffs decreased as targets were less flat and less visually distinct. Less flat targets change appearance more between the two camera viewpoints, and with inaccurate mast calibration, the expected hazcam target appearance is less accurate and finds its best match further from the correct target in the hazcam image. Less visually distinct targets are tracked with less accuracy because the tracker locks onto more distinct nearby, features. To the extent that incorrect calibration warps the expected appearance of the target, tracking a neighboring target does not provide a good prediction of the location of the original target.

There remain four tests where refined kinematic handoff failed. The final feature of test 1 was 55 pixels off. The feature is on flat ground and the predicted appearance seems correct, but the position selected by the tracker looks nothing like the expected appearance of the target. This is the same feature that provided inexplicable results for both pure and alternate kinematic handoff from near navcams. The final feature of test 2 was off by 65 pixels. It represents the same physical object as the erroneous feature in test 1 and exhibits all of the same confusing behavior, except that it did not confuse the earlier handoff methods. The final feature of test 3 was off by 16 pixels. The feature was not visually distinct, and when warped to generate an expected appearance, it became less distinct. Tracking found a good match 16 pixels away. It is not clear whether this should be classified as a success. A larger tracking window might have found the target more accurately by using more of the neighboring structure. This was the approach used in the original SCIP implementation, which matched at reduced resolution (effectively increasing the size of the feature window) to take advantage of additional information in the neighborhood, then refined the result by matching at normal resolution to improve localization. The final feature of test 5 tracked incorrectly due to what is surely a code bug. The target is toward the left of the joint field of view of the navcams, and thus near the left edge of the stereo data used to generate the expected hazcam target image. That image has black pixels, representing valid data, starting about 7 pixels to the left of the target. The tracker is supposed to ignore these black pixels, but the hazcam image, at the final coordinates chosen by the tracker, contains a light-to-dark transition exactly the same shape as the valid-to-invalid transition in the expected image. The fact that the erroneous match was only 18 pixels away is coincidental.

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.

E. Refined kinematic handoff from far navcams

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 25 tests, 17 successfully tracked the target into the hazcam while 8 failed. The successes found the target with average error of 1 pixels. This is slightly better than refinement from near navcams, which makes sense because the targets have similar appearance in the hazcam and far navcam images, so there is less opportunity to introduce error by warping the images. Six of the eight failures represented features whose appearance changed significantly as the feature moved into the rover shadow in the hazcam image. The features either were not distinct enough and evaporated as a result of being shaded, or their appearance changed enough that similar features elsewhere in the image looked more like the original target. These sorts of errors are unique to refined handoff from far navcams, as the other methods either take all images at once or do not rely on navcam and hazcam images having similar intensity patterns. Of the two remaining failures, one is inexplicable, as the target has similar appearance in the navcam, hazcam, and expected hazcam images. The other has only 8 pixel error and probably results from poor warping to generate the expected hazcam image, based on dramatically underestimating the amount of yaw experienced by the rover while approaching the target. The expected target appearance does not match the actual appearance, so the target drifts. Perhaps this should be considered inaccuracy rather than failure, though the magnitude of the error certainly makes it an outlier.

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 eight failures did occur in arced runs, but only in one case did warping from the different viewpoints made any difference.

F. 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. In fact, handoff error averaged 8 pixels, generally twice as accurate as pure kinematic handoff. The two notable instances where pure geometric handoff was worse than pure kinematic occur on two targets that both represent the same physical feature that inexplicably confused refined kinematic handoff. Pure geometric handoff succeeds in 24 tests and is unable to perform one test because the feature is too close to the edge of the image to generate a tracking window.

G. 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 59 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. Pure geometric handoff from far navcams succeeded for all 25 tests, though it did particularly poorly on the one feature that could not be tested in pure geometric handoff from near hazcams. Perhaps more care must be taken to handle the highly distorted edges of the hazcam images.

H. 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 21 tests, fails on 3, and cannot be performed on the one feature that is too close to the edge of the image. When it is successful, it generally recovers the target location accurate to 1 pixel, well within measurement accuracy. In all three failure cases, the geometry of the feature changed caused the appearance to change significantly between hazcam and navcam viewpoints, and new but incorrect hazcam features provided better matches to the original, navcam targets. In one case, a poor estimate of rover motion produced an inaccurate expected hazcam image. In the second case, the tracker accurately tracked a shadow on a rock, even though the shadow's shape changed with the new viewpoint. Unfortunately, the goal had been to track a feature inside the shadow, which moved relative to the shadow. Perhaps the coarse-to-fine strategy mentioned earlier would have tracked the correct point in the target window. The third case is a feature that had also confused refined kinematic handoff, being similar enough to the target that a little noise from viewpoint change and perhaps change in image intensity caused the tracker to switch to the wrong feature.

I. 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, as is readily visible in many of the expected-hazcam-appearance images. This method succeeded in only 2 of the 25 test cases. Those two successful cases were accurate to within 1 pixel. They corresponded to targets that, despite being draped incorrectly over the point cloud, happened to retain their shape well enough to produce a trackable feature. Perhaps visual odometry could remedy the pose error, the resulting draping error, the resulting error in expected target appearance, and the resulting error in handoff.

J. Mesh registration from near and far navcams

Mesh registration performed both well and consistently. Failures were traced to wholesale failures of stereo correlation (JPL stereo failed to produce a 3D model from the far navcam images in test 4), holes in the navcam stereo models around some selected target points, and presumably bad initial localization of the initial far navcam stereo models. Follow up tests to confirm the latter failure causes were not performed, however this explanation is consistent with the other results (the pure kinematic algorithm performed badly on the same images).

Of interest to note is that mesh registration reasonably assumes that rover roll and pitch are well characterized. This assumption appears to be violated in test run #5 and #6, leading to failure of in one case (Figure 5, #5).

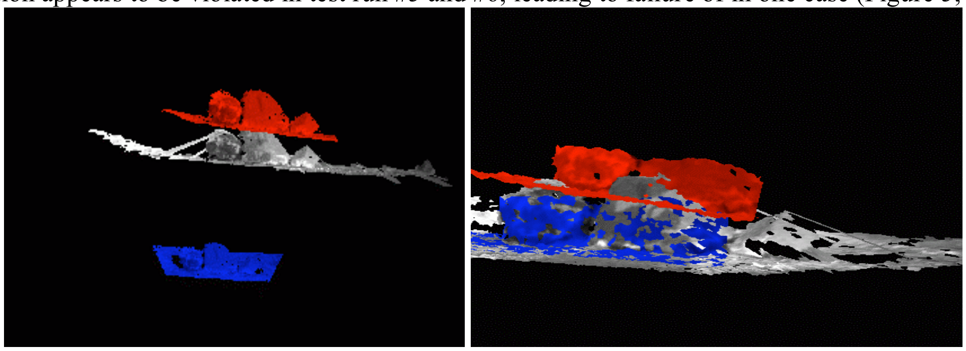


Figure 6 *Mesh registration results for far navcam tests 5 and 6.Red shows the initial far navcam models, aligned based on camera calibration and rover pose estimates. Grey is the hazcam model, and blue is the aligned navcam model after mesh registration.*

Failure of mesh registration was clearly detectable, either because no stereo data was available, or the alignment algorithm fails to converge to a physically plausible solution.

VI. Conclusions

A. Principal results

The most accurate, but unreliable, handoff algorithms were refined kinematic handoff from near or far navcams and refined geometric handoff from near navcams. These methods all recovered the correct handoff coordinates to within measurement error, approximately 2 pixels, in approximately half of the tests. The other half of the time, they failed completely. Refined geometric handoff from far navcams had similar accuracy when successful, but it was rarely successful. Currently, we cannot detect success or failure, making these methods of questionable use.

Mesh registration gives the next best results (4 pixels), without sacrificing reliability. The failures in mesh registration documented here are attributable to easily remedied causes: 1) no stereo correlation and 2) bad initial guess at the model alignments. Using autogain, or getting high dynamic range images would have solved the stereo failure in test #4. Seeding the mesh registration algorithm with better starting alignments, obtained either through better rover pose estimation, or with the results of a target tracking algorithm, as is done in the ARC approach. The drawback is that mesh registration requires the greatest computation.

All the above algorithms by themselves fall short of achieving the desired 1cm *total* placement error. Assuming the MER 0.5cm manipulator error budget, the combined tracking and handoff error budget is 2 pixels (assuming 1 pixel = 0.25cm error, as is the case in this experiment). Mesh registration, using templates from the initial pancam images would likely double this (Table 3). The refined kinematic/geometric algorithms do better (2 pixels), but their errors must be summed with similar target tracking errors, leading again to worse than required performance (Table 2)

Table 2 *JPL SCIP Target approach error budget (1 hazcam pixel = 0.25cm at 1m from target).* *<u>[Values] are unconfirmed estimates.</u>*

Pancam target tracking	Pancam-navcam handoff	Navcam target tracking	Navcam-Hazcam handoff	Hazcam target tracking	Total Error:

[1-2 pixels]	[1-2 pixels]	[1-2 pixels]	1-2 pixels (refined pure kinematic or pure geometric) 4-5 pixels (mesh registration)	[1-2 pixels]	[5 – 10 pixels] if refined pure kinematic/geometric handoff succeeds [8-14 pixels] using mesh registration
[1-2 pixels possible if the tracker can be made to work using a warped version of the original 2D pancam tracking template at each stage]					*[1-2 pixels] if tracker succeeds.*

Table 3 *ARC SCIP Target approach error budget. (1 Rocky 8 hazcam pixel = 0.25cm at 1m from target). [Values] are unconfirmed estimates.*

Target tracking	Pancam – Hazcam handoff	Total Error (= Handoff Error)
[3-6 pixels] using JPL affine tracker with intermediate pancam to navcam handoff 1.3–1.7cm using ARC keypoint tracker	4-5 pixels	4-5 pixels

One possible solution, suggested in Table 2 is to change the JPL tracker to always use a warped version of the original template image as the tracker template, rather than updating it with new navcam/hazcam images after each handoff. This presupposes that the tracker can be made robust to the additional image view changes involved.

B. Limitations

This work was intended to be a quick implementation / investigation of the refined geometic handoff technique described in [13]. It grew 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.

No pancam images, taken from 10m distant, were used in these tests. The mesh registration 10m pancam to 1m hazcam handoff accuracy quoted in Table 3 is inferred from the 3m (far) navcam to 1m hazcam handoff results. This is reasonable, given the 3x increased resolution of the pancams over the navcams, and supported by the small variation in the near navcam and far navcam mesh registration accuracies, and reported results of 1-2cm accuracy from earlier tests with ARC's K9 rover.

The work does not attempt to exhaustively sample the field of possible handoff algorithms. Most notably, it does not consider algorithms using fiducials in the scene, visible to both the navcams and hazcams (on the arm for example) to determine the 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.

The cameras on Rocky8 were not synchronized at the time of these tests. 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 in 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 reduces the reported handoff accuracy.

The pose estimation used during SCIP is not particularly accurate. Rover x,y,yaw are inferred by integrating commanded wheel rotations. Rover z is assumed constant; roll and pitch are directly measured from a suspect IMU, that clearly shows discrepancies (Figure 6). This unfairly penalizes the far-navcam handoff tests.

C. 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.

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.

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 were not reliable. In the 18 test cases in this work, and in previous work, there was 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.

Gain matching—Regulating camera gain so that the target has comparable intensity in each image might help the NCC tracker find the correct target more often, and it might eliminate the washed out images that occasionally fool some of the handoff algorithms.

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 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.

Acknowledgements

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

References

[1] Liam Pedersen 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),* Edmonton, 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.

[15] Baumgartner, E., R. Bonitz, J. Melko, L. Shiraishi, P. Leger, A. Trebi-Ollenu,”Mobile Manipulation for the Mars Exploration Rover,” in *IEEE Robotics and Automation Magazine,* Vol 13, No. 2, June 2006.