

# Rover Mast Calibration, Exact Camera Pointing, and Camera Handoff for Visual Target Tracking\*

Won S. Kim, Adnan I. Ansar, and Robert D. Steele

*Jet Propulsion Laboratory  
California Institute of Technology  
4800 Oak Grove Drive, Pasadena, CA 91109-8099, USA  
Won.S.Kim@jpl.nasa.gov*

**Abstract** – This paper presents three technical elements that we have developed to improve the accuracy of the visual target tracking approach system. An accurate yet quite simple method of rover mast calibration is achieved by using a total station, a camera calibration target, and four prism targets mounted on the rover. This method not only covers a large work volume for improved accuracy but also determines all camera models relative to the rover reference frame. The method was applied to Rocky8 rover mast calibration and yielded a 1.1-pixel rms residual error. Camera pointing requires inverse kinematic solutions for mast pan and tilt angles such that the target image appears right at the center of the camera image. Two issues were raised. Mast camera frames are in general not parallel to the masthead base frame. Further, the optical axis of the camera model in general does not pass through the center of the image. Despite these issues, we managed to derive non-iterative closed-form exact solutions, which were verified with a Matlab code. Finally, a purely geometric method for camera handoff has been developed by assuming stereo camera views of the target are available. Experimental test runs show less than 2 pixels error on high-resolution Navcam for Pancam-to-Navcam handoff, and less than 3.3 pixels error on lower-resolution Hazcam for Navcam-to-Hazcam handoff.

**Index Terms** - rover mast calibration, camera pointing, camera handoff, visual target tracking, target approach.

## I. INTRODUCTION

The baseline operation of the Mars Exploration Rover (MER'03) flight mission depicts the state of art technology for target approach and instrument placement on Mars. When a target rock is about 10 to 20 m away from the rover, the MER baseline operation requires 3 sols (Martian days) for the rover to place an instrument on the designated target. From panoramic images received from Mars, scientists select a target that might be 10 to 20 m away from the rover. Thereafter a waypoint is determined at about 2 to 4 m away from the target and sent to the rover. When the rover reaches the commanded waypoint next day, it takes images and sends them to Earth. Using these images, scientists and ground operators determine the rover stationary base or anchor position as the next waypoint where the target rock is within rover arm's reach. When the rover reaches this waypoint next day, it takes close-up target images. From these close-up images, scientists and ground operators determine rover arm motion commands with appropriate arm collision checking. The rover follows the commands

next day, placing an instrument on the target position of the rock. The above scenario takes 3 sols.

In future Martian surface operations such as Mars Science Laboratory (MSL'09), it is desirable to achieve the entire 10-m target approach and instrument placement in a single sol. If 8 to 10 sols are to be spent per rock with 3-sol baseline instrument placement, those numbers will be reduced to 6 to 8 sols per rock with the single-sol instrument placement capability, yielding 20% to 25% increase in science return.

The single-sol target approach and instrument placement is technically challenging. In particular, the operation must be fail-safe and reliable. Two major technologies to achieve single-sol target approach and instrument placement operations include 1) visual target tracking for approach and 2) rover-stereo-based manipulation to place an instrument. Related technologies were demonstrated earlier for some experimental conditions [1], [2], [3], [4]. Also there have been efforts to test and validate these technologies [5], [6]. However, fail-safe, reliable operations have not been demonstrated yet. Further enhancements and extensive experiments are necessary to produce fail-safe, reliable operations for target approach and instrument placement.

After an overview of a target approach system, this paper describes three technical elements that we developed to improve the accuracy of the visual target tracking approach system: 1) rover mast calibration, 2) exact camera pointing for active camera control, and 3) camera handoff.

## II. TARGET APPROACH SYSTEM

### A. Theoretical Calculations of Target Approach Accuracy

The target approach accuracy can be greatly improved by employing visual target tracking. To see the improvement, we performed theoretical computations of the target approach accuracy over a 10 m travel with and without a visual target tracker. If no target tracker is used, two main factors contributing to the target positioning error are the stereo range "sensing" error and the rover navigation error after 10-m travel. The stereo range error  $\Delta R$  is computed by

$$\Delta R = \frac{R^2}{f_p B} \Delta d, \quad (1)$$

where B is the stereo baseline, R is the range,  $\Delta d$  is stereo disparity error, and  $f_p$  is the camera's equivalent focal

---

\* This work was performed at the Jet Propulsion Laboratory, California Institute of technology, under a contract with the National Aeronautics and Space Administration.

length in pixels. The stereo disparity error  $\Delta d$  is assumed to be 1 pixel for the 3- $\sigma$  stereo range error [5]. The camera focal length  $f$  can be converted to pixels by

$$f_p = \frac{f}{\text{pixel size}}. \quad (2)$$

In Table I, we assume all cameras use a 1/3-inch CCD image sensor with a 1024×768 pixels resolution. Since the effective image size of the 1/3" CCD image sensor is approximately 4.8 mm × 3.6 mm, each CCD pixel is a square with the pixel size of 4.8/1024 = 3.6/768 = 0.00469 mm (manufacturer's specification of the actual pixel size should be very close to this value.) Further we consider three different cameras: Pancam, Navcam, and Hazcam. Their focal lengths are 16 mm, 6 mm, and 2.3 mm, respectively, while their stereo baselines are 30 cm, 20 cm, and 10 cm, respectively. These camera specifications are very close to those of the JPL Rocky8 Rover platform used in our tests except that Hazcam cameras currently have a lower resolution of 640×480. The specifications are also close to those of the MER with 1024×1024 pixels image resolution [7]. Based on the above relations, stereo range errors at 10-m distance are computed and listed in the second column of Table I.

TABLE I  
TARGET TRACKING ERROR WITHOUT AND WITH VISUAL TRACKING

Camera specifications (1/3" CCD image sensor)	Stereo range error (3 $\sigma$ ) at 10 m distance	Target approach error (3 $\sigma$ ) with 2% navigation error (no visual tracking)	Target approach error (3 $\sigma$ ) with ideal visual tracking and camera handoff
Pancam, 16 mm FOV=17° × 13° B=30 cm	9.7 cm	22.2 cm	1.5 cm
Navcam, 6 mm FOV=49° × 37° B=20 cm	38.8 cm	43.7 cm	3.9 cm
Hazcam, 2.3 mm FOV=113° × 86° B=10 cm	202.2 cm	203.2 cm	10.1 cm

If no visual target tracking is used, the target positioning error after 10-m travel to the target is the root-sum-square (RSS) of the stereo range error at 10 m and the rover navigation estimation error for 10 m travel.

$$\Delta R_{no\_tracking,10m} = \sqrt{\Delta R_{stereo,10m}^2 + \Delta R_{nav,10m}^2}. \quad (3)$$

If we assume the rover navigation error based on the wheel and visual odometers is roughly 2% of the rover travel distance, the navigation error  $\Delta R_{nav,10m}$  over 10 m travel is 20 cm. The computation results are listed in the third column of Table I. The target positioning errors without visual tracking are all more than 20 cm.

When a visual target tracker is employed, the target positioning error can be greatly reduced. Assuming the stereo-based manipulation is performed at 1-m distance, the positioning error of the target which is being tracked over camera images is approximately determined by the stereo range error at 1 m distance. Since a 1-pixel ( $\pm 0.5$ -pixel) image expands to a 10-pixel ( $\pm 5$ -pixel) image as the camera approaches the target from 10 m away to 1 m away, a 1-pixel target position accuracy at 10 m away can yield a  $\pm 5$ -

pixel accuracy in 1 m away. Since the target position cannot be estimated more accurately than  $\pm 5$  pixels at 1 m distance, the 3 $\sigma$  error stereo disparity error  $\Delta d = 5$  pixels at 1 m distance if there was no camera handoff. When the camera handoff takes place, the camera focal length is reduced from  $f_{s1}$  to  $f_{s2}$ . The corresponding stereo disparity error is also reduced by a factor of  $f_{s2}/f_{s1}$ , and thus the stereo range error is given by

$$\Delta R = \frac{R^2}{f_{s2} B_2} (\Delta d \frac{f_{s2}}{f_{s1}}) = \frac{R^2}{f_{s1} B_2} \Delta d, \quad (4)$$

where  $f_{s1}$  is the focal length of the initial camera used at 10 m away,  $B_2$  is the baseline of the final camera after handoff, and  $\Delta d = 5$  pixels for 3- $\sigma$  error value. This formula provides the theoretical target approach positioning error assuming ideal visual tracking with perfect camera handoff. The results are shown in the last column of Table I. The theoretical 3- $\sigma$  error of the target positioning is 1.5 cm with ideal visual tracking starting with Pancam and handing off to Navcam and then finally to Hazcam.

Table I clearly demonstrates that camera handoff is essential to attain high accuracy in target approach. Therefore, the following baseline operational scenario can be conceived that involves two camera handoffs.

1. Pancam for 4 m (from 10 m to 6 m)
2. Handoff from Pancam to Navcam
3. Navcam for 4m (from 6 m to 2 m)
4. Handoff from Navcam to Hazcam
5. Hazcam for 1m (from 2 m to 1 m)
6. Anchor rover and place instrument

Pancam to Navcam handoff is done at about 6 m from the target since the minimum stereo range for the Pancam is about 5 m. Navcam to Hazcam handoff is done at about 2 m from the target before target viewing angle from Navcam becomes too steep.

### B. 2-D/3-D Visual Target Tracking System

Fig. 1 shows a functional diagram of a target approach system consisting of the 2-D/3-D visual tracker and a camera handoff module. The 2-D/3-D tracker was implemented by Nesnas, Bajracharya, et al. [8]. It is essentially the 2-D tracker with active camera control. The 2-D tracker is a feature image matcher and does not require stereo camera views, while the active camera control does.

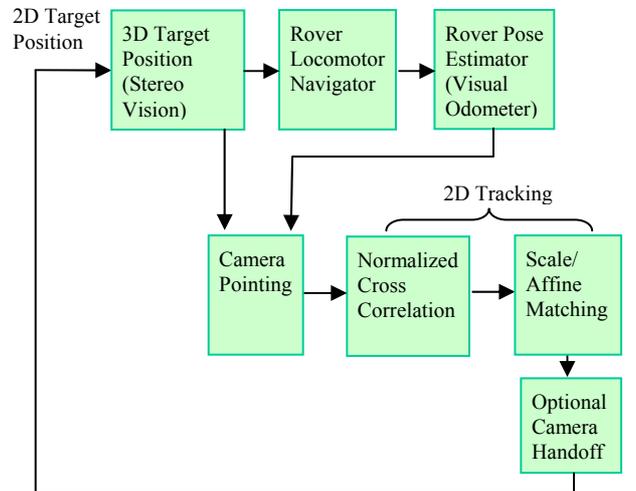


Fig. 1 2-D/3-D visual target tracking system.

Active camera control points the camera to the target each time when the rover moves to a new position so that the target image appears at the center of the camera image. This not only prevents the target image from moving out of the camera view but also reduces the 2-D tracker search area, enhancing the tracking reliability significantly. Pointing the camera to the target requires the knowledge of the target position in 3D space relative to the rover position. The rover pose estimator such as the visual odometer provides the rover pose estimate, while the triangulation of the target image points in stereo camera views provides the target position estimate in 3D space. Exact camera pointing requires accurate rover mast calibration.

### III. ROVER MAST CALIBRATION

#### A. Related Work on Camera Calibration

We describe camera calibration techniques first since our rover mast calibration method presented here requires camera calibrations beforehand. Camera calibration determines the camera model that defines the image formation geometry between 3-D coordinates of a point in the scene and its corresponding 2-D coordinates on the camera image. In order to avoid 3-D measurements of target positions, many researchers simply used one calibration target fixed at one position for camera calibration. It has two problems. First, the camera model is given relative to the fixed calibration target frame not to the rover frame. This is problematic, since the target position relative to the rover is not accurately known. Second, it does not cover adequate work volume for measurements. Since it is impractical to build a very large accurate calibration target, Shih et al. [9] used a 9×9 dot calibration target on a linear translation stage, whose position was measured by a CMM (coordinate measurement machine). This encompasses one strip section only. Davis and Chen [10] used a bright LED light and a 3-D tracker to measure 3-D positions of large work volume. Although a 3-D tracker following a LED light would not be very accurate, their experimental results demonstrated the importance of covering a large work volume. In this paper we present an accurate yet quite simple method to calibrate the rover mast using a total station (surveying equipment that measures 3-D position), a camera calibration target, and four prism targets mounted on the rover. This method not only covers a large work volume but also determines camera calibration parameters relative to the rover reference frame fixed to the rover.

#### B. Camera Calibrations Relative to Rover Frame

Fig. 2 shows a typical setup to collect data for camera calibration. The camera calibration target is made of a light, inexpensive gator-foam board with 10×10 dots pattern. Three reflective-tape targets are attached at three corners of the target board. A Leica TCRA 1103 total station (surveying equipment) with 2-mm metrology accuracy is used to measure 3-D positions of reflective tape targets attached on the calibration target board. The target stand is used to facilitate positioning the calibration target board, which can be raised or lowered at different tilt angles. Table II lists the rover body tilt, mast tilt, and calibration target

positions used to calibrate the three camera pairs of Rocky8 rover. In Pancam and Navcam calibrations, the mast is tilted downward since the calibration target cannot be placed too high. In Hazcam calibration, since Hazcam is tilted downward and near the surface, the rover needs to be tilted backward to allow adequate work volume for calibration target positions. For each camera calibration, the calibration target is moved to 7 or 8 positions [5]. At each position, the camera images are taken and 3-D target positions are measured. These collected data are used to calibrate the camera. There are a number of camera calibration software packages. We used the JPL camera calibration software that generates the CAHVOR camera model [11].

There is, however, a particularly important procedure in our calibration method. All cameras must be calibrated in the same reference frame to attain good accuracy for camera handoff. Four 360° prism targets on the poles mounted on the rover were used for this purpose (see Fig. 2). Accurate positions of the four prism positions relative to the rover reference frame must be given. If initial prism positions are not accurate, they need to be refined iteratively at different total station locations. Since the prisms are only 40 cm to 60 cm apart, extra care is needed to determine the rover reference frame. It is best not to move the total station when calibrating cameras. Experimental tests of 40 position measurements showed that the maximum position error relative to the rover reference frame was about 6 mm for about 8-m away targets. This corresponds to 0.3 mm error for 40-cm apart prisms, and to approximately 0.00075 radians or 0.043° error in defining the rover reference frame. When the total station is moved, however, the position error increased significantly as shown in Table III.

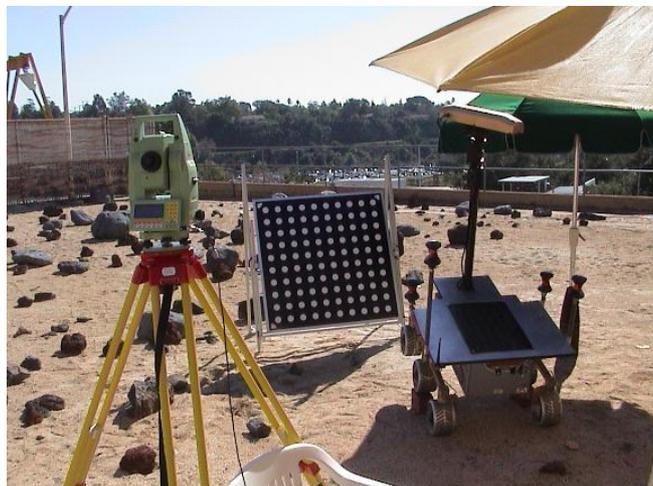


Fig. 2 A typical setup to collect camera calibration data: total station (left), calibration target and its stand (middle), Rocky8 rover with a pan/tilt mast and four prisms (right). Courtesy of Max Bajracharya.

TABLE II  
CAMERA CALIBRATION POSITIONS

	Approximate rover tilt	Approximate mast tilt	Calibration target positions
Pancam	0°	-5°	3 or 4 positions at ~5 m 4 positions at ~10 m
Navcam	0°	-10°	3 or 4 positions at ~3 m 4 positions at ~6 m
Hazcam	-30°	0°	3 or 4 positions at ~1.5 m 4 positions at ~3 m

TABLE III  
METROLOGY ERROR TO MEASURE RELATIVE TO ROVER FRAME

Total station positions	Maximum error for 6-m away targets	Corresponding error for 40-cm apart prisms	Worst error in defining rover reference frame
Same positions	6 mm	0.3 mm	0.00075 rad. 0.043°
Different pos.; facing	34 mm	1.7 mm	0.00425 rad. 0.24°
Different pos.; fixed	109 mm	5.5 mm	0.0136 rad. 0.78°

### C. Rover Mast Calibration

Our mast calibration method is a simple extension of the camera calibration previously described. Collect camera images and 3-D target positions at several different calibration target positions with different mast pan and tilt angles. The key idea here is again to get 3-D metrology data relative to the same rover reference frame used in camera calibrations. This is critical to attain accurate camera handoff, for example, from mast-mounted Navcam to body-mounted Hazcam. To minimize the discrepancy between rover reference frame definitions, it is best to keep the same total station position for the mast and camera calibrations. In other words, continue data collection for mast calibration after the camera calibrations without moving the rover and the total station.

Rover mast calibration requires definitions of the rover mast kinematics. In Rocky8 rover, the Pancam/Navcam masthead is mounted on a vertical mast with a 2-dof (degrees of freedom) pan-tilt unit (see Fig. 2). Fig. 3 defines coordinate frames for rover mast kinematics. From Fig. 3, the camera frame is related to the world reference frame by

$$T_{w-camera} = T_{rover} * T_{mast} * T_{masthead} * T_{camera}, \quad (5)$$

where

$$\begin{aligned} T_{rover} &= \text{rover frame relative to the world frame,} \\ T_{mast} &= \text{mast pan-tilt base frame relative to rover,} \\ T_{masthead} &= \text{masthead base frame relative to mast,} \\ T_{camera} &= \text{camera frame relative to masthead base.} \end{aligned}$$

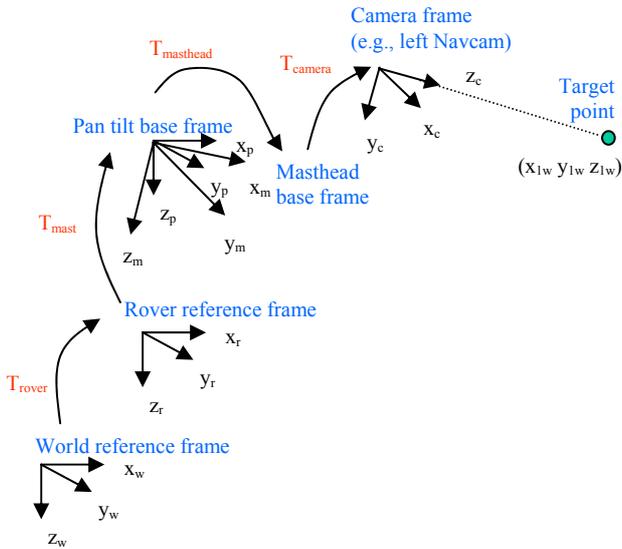


Fig. 3 Coordinate frame definitions for rover mast kinematics.

The rover frame is defined such that the z-axis is down, the x-axis is forward, and the y-axis is to the right, while the camera frame is defined such that the z-axis is forward, the x-axis is to the right, and, the y-axis is down. In mast calibration, we are interested in the camera frame relative to the rover reference frame.

$$T_{r-camera} = T_{mast} * T_{masthead} * T_{camera}. \quad (6)$$

The transform from the rover reference frame to the mast pan-tilt base frame can be described by three translation and two rotation parameters:

$$\begin{aligned} T_{mast} &= Trans(t_{xm}, t_{ym}, t_{zm}) * Rot(x, \theta_{xm}) \\ &\quad * Rot(y, \theta_{ym}). \end{aligned} \quad (7)$$

The rotation parameter about the z-axis is not needed since the subsequent transform  $T_{masthead}$  starts with the z-axis rotation.  $T_{masthead}$  is a function of pan and tilt angles:

$$\begin{aligned} T_{masthead} &= Rot(z, pan + pan\_offset) \\ &\quad * Rot(y, tilt + tilt\_offset), \end{aligned} \quad (8)$$

where the  $pan\_offset$  and  $tilt\_offset$  parameters are determined by the mast calibration.

Input data to mast calibration are 3-D points, their corresponding 2-D image points, pan, tilt angles, and camera models as well as pan and tilt angles at the time of the camera calibrations. Given input data, the mast calibration applies a nonlinear least-squares method to determine seven parameters:  $t_{xm}$ ,  $t_{ym}$ ,  $t_{zm}$ ,  $\theta_{xm}$ ,  $\theta_{ym}$ ,  $pan\_offset$ , and  $tilt\_offset$ . It, however, turns out that  $tilt\_offset$  is a free variable that can be set to any value mathematically. Therefore, only six parameters are determined by the nonlinear least-squares method with  $tilt\_offset = 0$ . Thereafter,  $tilt\_offset$  is adjusted such that the average of the camera optical axis vector of all four mast cameras lies on the plane parallel to the xy-plane of the masthead base frame.

The nonlinear squares method determines the mast calibration parameters  $\psi^*$  that minimize the mean square metric of the errors between the projected image points of 3-D points and their corresponding 2-D image points.

$$\psi^* = \arg \min_{\psi} \|C(\psi, K_o, pan_{K_o}, tilt_{K_o}, pan, tilt, X_{3d}) - X_{2d}\|, \quad (9)$$

where  $\psi$  is the six-parameter vector to solve

$$\psi = [t_{xm} \ t_{ym} \ t_{zm} \ \theta_{xm} \ \theta_{ym} \ pan\_offset]. \quad (10)$$

$X_{3d}$  and  $X_{2d}$  are vectors of 3-D points and the corresponding 2-D points.  $K_o$  is the camera model for each camera, obtained by the camera calibration procedure described previously. Since the camera calibration is relative to the rover frame, the pan and tilt angles  $pan_{K_o}$  and  $tilt_{K_o}$  at the time of camera calibration for each camera must be provided. The  $C$  function projects 3-D points onto the image plane, which requires the camera model relative to the rover frame,  $K_{rover}$ , for given pan and tilt angles.

To facilitate computing camera models, we define  $K_{masthead}$  as the camera model relative to the masthead base frame.  $K_{masthead}$  is fixed regardless of pan and tilt angles. Assuming we use CAHVOR camera models, we can compute  $K_{rover}$  from  $K_{masthead}$  for given pan and tilt angles.

Fig. 3 and (6) illustrate that  $T_{camera}$  is associated with  $K_{masthead}$ , while  $T_{r-camera}$  is with  $K_{rover}$ . It can be shown that

$$K_{rover} = transform\_K(K_{masthead}, T_{mast} * T_{masthead}), \quad (11)$$

where  $transform\_K()$  transforms the CAHVOR camera model where the second parameter  $T_{mast} * T_{masthead}$  defines the amount of rotation and translation. The  $transform\_K()$  rotates and translates the camera center position vector  $\mathbf{C}$ , while only rotates the  $\mathbf{A}$ ,  $\mathbf{H}$ ,  $\mathbf{V}$ , and  $\mathbf{O}$  vectors. It does not change the nonlinear parameter vector  $\mathbf{R}$ . Similarly, it can be shown that

$$K_{masthead} = transform\_K(K_o, (T_{mast} * T_{masthead-Ko})^{-1}), \quad (12)$$

where  $T_{masthead-Ko}$  is  $T_{masthead}$  with  $pan = pan_{Ko}$  and  $tilt = tilt_{Ko}$ . Thus, (12) followed by (11) computes  $K_{rover}$  from the given input parameters of (9) for each nonlinear least-squares iteration. The mast calibration algorithm determines the seven mast calibration parameters and also generates the CAHVOR models relative to the masthead, which are fixed regardless of pan and tilt angles. The user and other application software will then use (11) to compute the CAHVOR models relative to the rover for given pan and tilt angles.

We implemented the above mast calibration algorithm in Matlab code and tested with the Rocky8 pan/tilt mast. Initially we used a total of 18 calibration target positions, and observed that an adequate number of target positions was about 5 to 8. Fig. 4 shows a final set of 8 target positions selected: 3 outer positions from 5 to 10 m for Pancam and 5 inner positions from 1 to 5 m for Navcam. The pan angle ranged from -1 to 1 radian, and the tilt angle was from -0.1 to -1 radian.

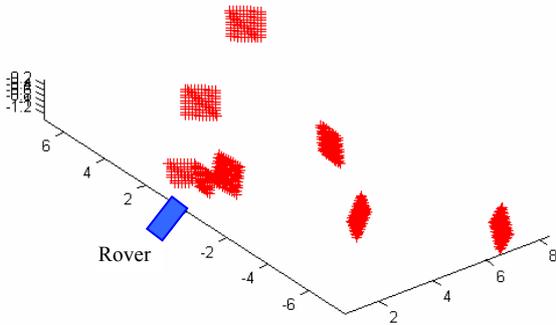


Fig. 4 Calibration target positions used for mast calibration.

Table IV lists the seven mast calibration parameters determined by the mast calibration Matlab code. The mast was not perfectly aligned with the z-axis of the rover reference frame; it tilted left by 0.0088 radians (0.50°) and forward by 0.0181 radians (1.04°). Table V lists the Pancam and Navcam 2-D residual rms errors of the mast calibration result. For the full seven-parameter mast calibration, the maximum residual rms error was 1.13 pixels. However, if the mast is assumed to be orthogonal or perfectly aligned with the z-axis of the rover frame with  $\theta_{xm} = \theta_{ym} = 0$ , the maximum residual rms error was increased dramatically to 22.29 pixels. Thus in the Rocky8 mast calibration, it is important to use full 7 parameters to attain high accuracy.

TABLE IV  
MAST CALIBRATION PARAMETERS IN METERS AND RADIAN

<i>mast</i> <i>t<sub>xm</sub></i>	<i>mast</i> <i>t<sub>ym</sub></i>	<i>mast</i> <i>t<sub>zm</sub></i>	<i>mast</i> <i>θ<sub>xm</sub></i>	<i>mast</i> <i>θ<sub>ym</sub></i>	<i>pan</i> <i>offset</i>	<i>tilt</i> <i>offset</i>
0.3443	-0.0043	-1.4572	-0.0088	-0.0181	0.0027	0.0178

TABLE V  
MAST CALIBRATION 2-D RESIDUAL RMS ERRORS IN PIXELS

	Pancam residual error		Navcam residual error	
	x	y	x	y
7 parameters	0.854	1.102	1.129	1.002
5 parameters	2.871	22.294	2.784	7.934

#### IV. EXACT CAMERA POINTING

Visual target tracking over a 10-m travel needs active camera pointing so that the target image to track is in the camera view. Further it is best to keep the target image at the center of the camera image so that the visual tracking does not need to search a large area. Smaller search area improves the tracking reliability. For given rover pose and target position estimates, best accuracy of camera pointing is achieved by using the exact inverse kinematic solutions for the pan and tilt angles such that the target image appears right at the center of the selected mast camera image. There are two issues to cope with to obtain the exact inverse kinematic solution: 1) the axes of the mast camera frames are in general not parallel to the axes of the masthead base frame, and 2) the optical axis (vector  $\mathbf{A}$ ) of the CAHVOR camera model in general does not pass through the center of the image. We managed to come up with a non-iterative closed-form exact solution, which is presented here.

From (5) and Fig. 3, the world frame coordinates of the target point are related to the camera frame coordinates by

$$\begin{bmatrix} x_{1w} \\ y_{1w} \\ z_{1w} \\ 1 \end{bmatrix} = T_{w-camera} \begin{bmatrix} x_{1c} \\ y_{1c} \\ z_{1c} \\ 1 \end{bmatrix}. \quad (13)$$

Since  $T_{rover}$  is given by the rover pose estimator and  $T_{masthead}$  are given by the mast calibration, the target point can be converted from the world coordinates ( $x_{1w} y_{1w} z_{1w}$ ) to mast frame coordinates ( $x_{1p} y_{1p} z_{1p}$ ). From (5) and (13),

$$\begin{bmatrix} x_{1p} \\ y_{1p} \\ z_{1p} \\ 1 \end{bmatrix} = T_{masthead} * T_{camera} \begin{bmatrix} x_{1c} \\ y_{1c} \\ z_{1c} \\ 1 \end{bmatrix}. \quad (14)$$

If the optical axis of the camera model passes through the center of the camera image, the target point appears at the center of the camera image when

$$x_{1c} = y_{1c} = 0. \quad (15)$$

Since the optical axis of the camera model in general does not pass through the center of the camera image, we need to compensate for the offset by rotating the camera frame such that the constraint of (15) let the target point re-appear at the center of the camera image.  $T_{camera}$  is related to the CAHVOR camera model of  $K_{masthead}$  by

$$T_{camera} = \begin{bmatrix} \mathbf{H}' & \mathbf{V}' & \mathbf{A} & \mathbf{C} \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (16)$$

where  $\mathbf{C}$  is the camera center position vector, and  $\mathbf{A}$  is the camera axis vector.  $\mathbf{H}$  and  $\mathbf{V}$  are horizontal and vertical information vectors, which are not in general mutually orthogonal, while  $\mathbf{H}'$ ,  $\mathbf{V}'$ , and  $\mathbf{A}$  are mutually orthogonal unit vectors along the x, y, and z axis, respectively. In general,  $\mathbf{H}'$  and  $\mathbf{V}'$  can be extracted from the CAHVOR model.

$$\mathbf{H} = h_s \mathbf{H}' + h_c \mathbf{A}, \quad (17)$$

$$\mathbf{V} = v_s \mathbf{V}' + h_c \mathbf{A}, \quad (18)$$

Parameters  $h_s$  and  $v_s$  are horizontal and vertical focal lengths expressed in pixels. The optical axis vector  $\mathbf{A}$  is perpendicular to the image plane, and passes through the image at the image coordinates  $(h_c, v_c)$  expressed in pixels. The image coordinates of the upper left corner is  $(0, 0)$ , and the image center is at  $(h_{max}/2, v_{max}/2)$ , where  $h_{max}$  and  $v_{max}$  are the image width and height. The amounts of horizontal and vertical rotational compensation required are

$$\theta_h = ATAN\left(\frac{h_{max}}{2} - h_c, h_s\right), \quad (19)$$

$$\theta_v = ATAN\left(\frac{v_{max}}{2} - v_c, v_s\right). \quad (20)$$

The transform needed for offset correction is given by

$$T_{offset} = Rot([- \theta_v, \theta_h, 0]), \quad (21)$$

where the vector  $[- \theta_v, \theta_h, 0]$  defines the rotation with the angle by its norm and the rotation axis by its unit vector. By inserting the offset correction transform into (14) and applying constraint (15),

$$\begin{bmatrix} x_{1p} \\ y_{1p} \\ z_{1p} \\ 1 \end{bmatrix} = T_{masthead} * T_{camera} * T_{offset} \begin{bmatrix} 0 \\ 0 \\ z_{1c} \\ 1 \end{bmatrix}. \quad (22)$$

From (22), we were able to derive closed-form exact kinematic solutions for pan and tilt angles such that the target image appears at the center of the camera image. We verified with the Matlab code that the solutions were within sub-pixel accuracy. Details are, however, omitted due to space limit.

## V. CAMERA HANDOFF

### A. Purely Geometric Camera Handoff

As described in Section II.A, camera handoff is essential to attain high accuracy in target approach. In an example baseline operation scenario, the target is initially tracked on Pancam images. On the way, it is switched to track on Navcam images, and finally switched to track on Hazcam images. Three techniques can be considered for camera handoff: 1) purely geometric computations, 2) 2-D image matching, and 3) 3-D stereo range matching. The purely geometric method can provide an initial estimate of the target image location for the next camera, while the image and range matching can potentially provide refinement of the estimate.

We implemented a Matlab test code for the purely geometric method, assuming stereo camera views are

available. The computational procedure is straightforward as follows.

1. Compute the rays from camera centers to target image positions for current left and right cameras.
2. Compute the 3-D target position as the intersection or the midpoint of the normal between the two rays.
3. Reproject the 3-D target position onto each of the next left and right cameras.

The camera handoff accuracy of the purely geometric method is directly affected by the accuracy of the mast and camera calibrations.

### B. Experimental Tests of Purely Geometric Handoff

As a way to measure the camera handoff accuracy, we came up with an idea of using bricks with reflective-tape targets attached at their four corners. This method provides two merits: 1) it enables to measure accurate 3-D positions of reflective-tape targets with a total station, and 2) it reduces ambiguity in human mouse-click entry of corresponding target points in four cameras (two sets of left and right cameras). In experimental tests, we placed bricks at about 6 m away for Pancam-to-Navcam handoffs and at about 2 m away for Navcam-to-Hazcam handoffs. We then measured reflective-tape target positions, recorded mast pan and tilt angles, and collected camera images. After the data collection, we entered the center positions of the reflective-tape target images manually by mouse-clicking. Since the mouse-click entry accepts only integers, it inherently has  $\pm 1$  pixel quantization error in target image position data.

A typical test run of a Pancam-to-Navcam handoff is shown in Fig. 5. Given human-entered target image points in Pancam stereo cameras, the Matlab test code computes the corresponding target image positions in Navcam stereo cameras by the purely geometric method. The cross marks

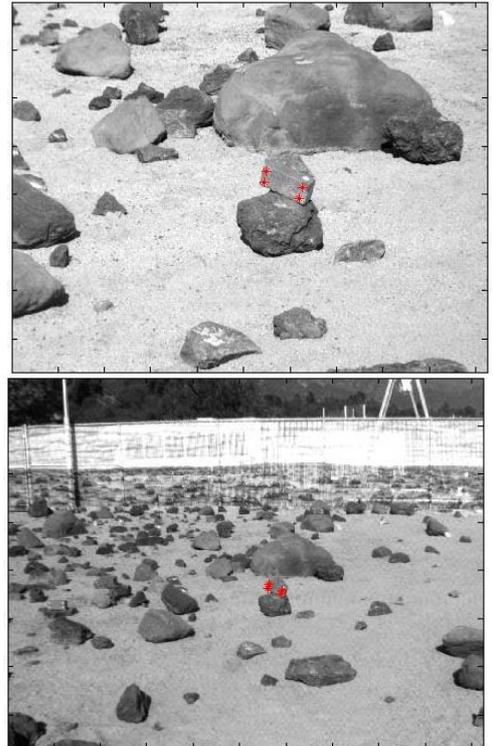


Fig. 5 Camera handoff from Pancam (left) to Navcam (right).

overlaid on the Pancam image in Fig.5 are human-entered target image positions, while those on the Navcam image are corresponding target positions determined by the purely geometric method. The difference between the computed and human-entered target positions for the Navcam image is the camera handoff error. The maximum error for four data points was 2.03 pixels, but there was a bias of about 1.5 pixels in the error data. The number is somewhat larger than expected, considering that the Pancam-to-Navcam handoff is not affected by the mast calibration error at all. One major source of the error appears to be the discrepancy in defining the rover reference frame during Pancam and Navcam camera calibrations. This would result in a constant bias in error. Inaccuracy in human data entry definitely contributes to the exaggeration of the handoff error. The Matlab test code also computed the stereo down-range error along the line-of-sight range direction and stereo cross-range error orthogonal to the line-of-sight direction. The targets were about 6.5 m away, and their maximum down-range and cross-range errors among 4 target points were about 30 mm and 2 mm, respectively.

A typical test run of a Navcam-to-Hazcam handoff is shown in Fig. 6. The cross marks overlaid on the Navcam image in Fig. 6 are human-entered target image positions, while those on the Hazcam image are corresponding target positions determined by the purely geometric method. The maximum camera handoff error for eight data points was 3.3 pixels. Since the Hazcam in the current Rocky8 set-up is a lower resolution of 640×480, the equivalent handoff error for a 1024×768 resolution is  $3.3 \times 1024 / 640 = 5.3$  pixels. The targets were about 2.0 m away, and their maximum down-range and cross-range errors among 8 target points were about 22 mm and 6 mm, respectively.

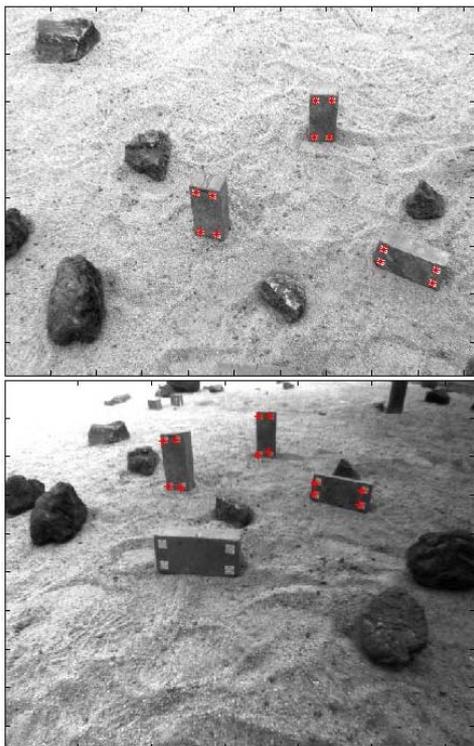


Fig. 6 Camera handoff from Navcam (left) to Hazcam (right).

The Navcam-to-Hazcam handoff error was larger than the Pancam-to-Navcam handoff error. This is because the mast calibration error directly affects Navcam-to-Hazcam handoff but not Pancam-to-Navcam. Further, there is usually a significant difference in viewing directions between Navcam and Hazcam during the Navcam-to-Hazcam handoff. Fig. 7 illustrates how the major-axis of the Navcam stereo range error ellipsoid is projected onto the Hazcam image. The projection length of the major-axis of the error ellipsoid onto the Hazcam image is approximately proportional to  $\sin \theta$ , where  $\theta$  is the angle between Navcam and Hazcam optical axes (viewing directions). As the  $\theta$  angle increases towards  $90^\circ$ , more error is propagated to the Hazcam.

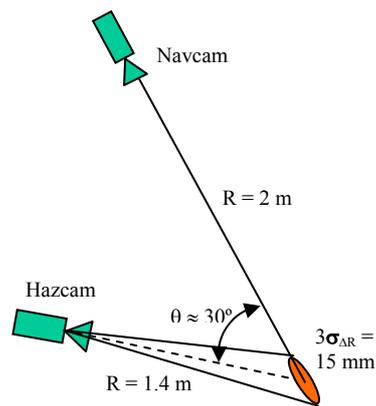


Fig. 7 Propagation of the Navcam stereo range error ellipsoid to Hazcam.

### C. Handoff Refinement

Since the Navcam-to-Hazcam handoff has error on the order of several pixels, we refine the pure geometric result using a simple correlation based technique. We first generate a template for correlation matching by warping a window centered at the target point in the Navcam image onto the Hazcam image. The registration between pixels needed for the warp is accomplished by projecting Hazcam stereo results onto the Navcam according to the camera models generated during the handoff. Fig. 8 illustrates the concept using the whole Navcam image rather than a small window for better visualization. We then run a 2-D normalized cross-correlation algorithm on the Hazcam image using the warped template. The starting point for the correlator is given by the geometric handoff of the Navcam target point.

Using the image in Fig. 8, we hand selected a pair of matched points (assumed error  $\sim 1$  pixel) in the Navcam and

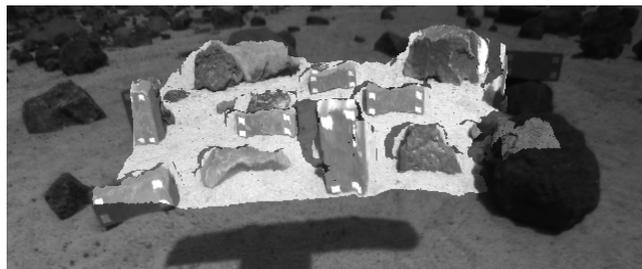


Fig. 8 Navcam image warped and superimposed (highlight) onto Hazcam image using camera models generated during handoff

Hazcam and then perturbed the match by 5-10 pixels in Hazcam image over multiple trials. We then ran the correlation step with a 21 x 21 template as well as a 21 x 21 search window. In each case, the true match point was recovered.

## VI. CONCLUSION

We developed three technical elements that improve the accuracy of the visual target tracking approach system: 1) rover mast calibration, 2) exact camera pointing, and 3) camera handoff. Algorithms were implemented in Matlab code, and experimental test results were presented. Test and validation of the entire 2-D/3-D visual target tracking system is on-going.

## ACKNOWLEDGMENT

This work is supported by NASA Mars Science Laboratory (MSL'09) Focused Technology Program. The authors would like to thank Max Bajrachrya, Issa Nesnas, and Richard Volpe for stimulating discussions and advice.

## REFERENCES

- [1] I. Nesnas and M. Maimone and H. Das, "Autonomous Vision-Based Manipulation from a Rover Platform," IEEE Symp. on Computational Intelligence in Robotics & Automation, 1999.
- [2] Huntsberger, T., H. Aghazarian, Y. Cheng, E.T. Baumgartner, E. Tunstel, C. Leger, A. Trebi-Ollennu, and P.S. Schenker, "Rover Autonomy for Long Range Navigation and Science Data Acquisition on Planetary Surfaces," in proc. IEEE Int. Conf. on Robotics and Automation, Washington, D.C., May 2002.
- [3] Pederson, L., M. Bualat, D.E. Smith, R. Washington, "Integrated Demonstration of Instrument Placement, Robust Execution and Contingent Planning", in International Symposium on Artificial Intelligence and Robotics in Space (i-SAIRAS) 2003, Nara, Japan, 2003.
- [4] M. Deans, C. Kunz, R. Sargent, L. Pederson, "Terrain Model Registration for Single Cycle Instrument Placement," Proc. i-SAIRAS, 2003.
- [5] W. Kim, R. Steinke, R. Steele, and A. Ansar, *Camera Calibration and Stereo Vision Technology Validation Report*, JPL D-27015, Jan. 2004.
- [6] W. Kim, R. Steinke, R. Steele, *2-D Target Tracking Technology Validation Report*, JPL D-28523, Apr. 2004.
- [7] Eisenman, Liebe, Maimone, Schwochert, and Willson, "Mars Exploration Rover Engineering Cameras," SPIE Remote Sensing conference proceedings, Toulouse, France, Sept. 2001.
- [8] I. Nesnas, M. Bajracharya, R. Madison, E. Bandari, C. Kunz, M. Deans, M. Bualat, "Visual Target Tracking for Rover-based Planetary Exploration," IEEE Aerospace Conference, Big Sky, Montana, 2004.
- [9] S. W. Shih, Y. P. Hung, and W. S. Lin, "Calibration of an active binocular head," IEEE Transactions on Systems, Man & Cybernetics, Part A (Systems & Humans), vol. 28, no. 4, pp. 426-442, 1998.
- [10] J. Davis and X. Chen, "Calibrating pan-tilt cameras in wide-area surveillance networks," IEEE International Conference on Computer Vision, 2003.
- [11] D. Gennery, *Camera Calibration including Lens Distortion*, JPL Technical Report, D-8580, Jet Propulsion Laboratory, Pasadena, CA, 1991.