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 for single-sol approach-and-instrument placement in future Mars rover missions. An accurate, straightforward method of rover mast calibration is achieved by using a total station, a camera calibration target, and four prism targets mounted on the rover. 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 Matlab routines. Actual camera pointing experiments over 50 random target image points yielded less than 1.3-pixel rms pointing error. Finally, a purely geometric method for camera handoff using stereo views of the target has been developed. Experimental test runs show less than 2.5 pixels error on high-resolution Navcam for Pancam-to-Navcam handoff, and less than 4 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 represents 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, MER baseline operations require 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 the 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 at least 3 sols.

In future Martian surface operations such as Mars Science Laboratory (MSL'09) mission, 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 the 3-sol baseline instrument placement capability, 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, and further enhancements and extensive experiments are necessary.

After an overview of a target approach system, this paper describes three technical elements that we developed to improve the robustness 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*

Target approach accuracy can be improved by using 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 pose estimation error after 10-m travel. The stereo range error ΔR is computed by

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

where B is the stereo baseline, R is the range, Δd is stereo disparity error, and f_p is the camera's equivalent focal length in pixels. The stereo disparity error Δd is assumed to

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

be 1 pixel for the 3-σ stereo range error [5]. The camera focal length f can be converted to pixels by

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

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 is very close to this value). Further we consider three different cameras: mast-mounted Pancam and Navcam, and body-mounted 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 close to those of the JPL Rocky8 Rover used in our tests except that Hazcam cameras have a lower resolution of 640×480. The specifications are also close to those of 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 (3σ) at 10 m distance	Target approach error (3σ) with 2% navigation error (no visual tracking)	Target approach error (3σ) 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 second data 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 reduced greatly. 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 (±0.5-pixel) image expands to a 10-pixel (±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 ±5-pixel accuracy at 1 m away. Since the target position cannot be estimated more accurately than ±5 pixels at 1 m distance, the 3σ stereo disparity error Δd = 5 pixels at 1 m distance without 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 Δd = 5 pixels for 3-σ 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-σ 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 the target viewing angle from the 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.

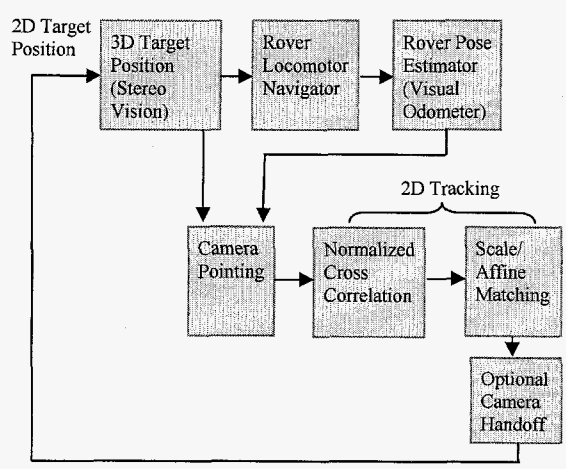

Fig. 1 2-D/3-D visual target tracking system.

Active camera control points the camera to the target each time 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. A rover pose estimator such as the visual odometer provides a 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 the rover mast calibration method presented here requires camera calibration 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 in the camera image. Vision researchers have traditionally used fiducials with known 3-D geometry in some external reference frame. In our case, these points must encompass an adequate work volume of the camera for rover applications. In addition, an accurate transformation between the global coordinate frame and the rover frame must be recovered. 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 covers one strip of the scene only. Davis and Chen [10] used a bright LED light and a 3-D tracker to measure 3-D positions of a large work volume. Although a 3-D tracker following an LED 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 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 a 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 calibration target is made of a light, inexpensive gator-foam board with a 10×10 dot pattern. Three reflective-tape targets are attached to three corners of the target board. A Leica TCRA 1103 total station (surveying instrument) with 2-mm metrology accuracy is used to measure 3-D positions of reflective tape targets attached to the calibration target board. A 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 the 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 the 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 with the total station. These collected data are used to calibrate the camera.

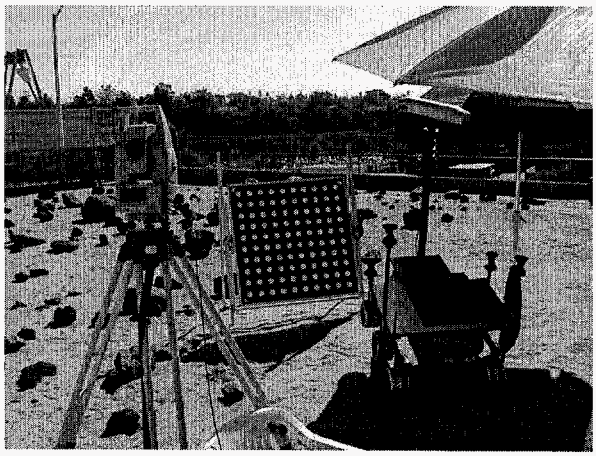

Fig. 2 A typical setup to collect camera calibration data: total station (left), calibration target and its stand (middle), Rocky8 rover with a pan/tilt mast and four prisms mounted on poles (right).

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

It is essential that all cameras be calibrated in the same reference frame to establish accurate geometric relations between the cameras. Four 360° prism targets on 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 the CAD model of the prism positions is inaccurate, it needs to be refined iteratively at different total station locations. Since the prisms are only 40 cm to 60 cm apart, orientation error is magnified at a distance. By noting that the prism target has a 2-*mm* metrology error, this corresponds to a 0.005 radians or 0.3 degrees error in defining the rover reference frame. Thus it is best to use the same prism measurements if possible, for instance, for all Pancam and Navcam mast cameras. For the Hazcam cameras, however, the rover needs to be tilted backward (Table II), and thus the prism target positions must be measured separately. It should be noted that separate prism measurements can introduce the systematic bias error of up to 0.005 radians or 0.3 degrees error in defining rover reference frames.

For camera calibration, we used the JPL camera calibration software that generates the CAHVOR camera

model [11]. This is because JPL flight missions including Mars Exploration Rover (MER) use the CAHVOR camera model. The techniques described in this paper are, however, applicable to any camera model. The CAHVOR model parameterizes the imaging geometry using six 3-dimensional vectors indicated by the six letters in the model name; **C** is the camera center position vector, **A** is the camera optical axis vector, **H** and **V** are the horizontal and vertical information vectors, and **O** and **R** are for lens distortion. The **A** and **O** vectors are unit length, so that the parameter space is 16-dimensional. Given a 3-D point **P** represented in the global coordinate frame of the model, **P** is mapped to **P'** via a polynomial warping of space symmetric about the vector **O** and parameterized by coefficients in **R**. Let **H**, **V** and **A** be written as row vectors. Then,

$$\mathbf{X} = \begin{bmatrix} \mathbf{H} \\ \mathbf{V} \\ \mathbf{A} \end{bmatrix} (\mathbf{P'} - \mathbf{C}) . \tag{5}$$

X is the homogeneous image coordinates of the projection of **P** so that the resulting pixel coordinates are $(X_1/X_3, X_2/X_3)$. The linear component of this mapping is simply a re-parameterization of the standard (sometimes called Tsai [12]) camera model and a 6-DOF Euclidean motion. The non-linear component of the CAHVOR model differs from the standard model by warping 3-D coordinates rather than the image plane. As a practical matter, this affords a better modeling of the barrel axis of the lens (the **O** vector) and is more flexible than purely radial image distortion models. However, it does not capture the tangential distortion associated with lens element misalignment [13]. In practice, the models are equally effective.

The camera calibration described above determines the camera model relative to the rover frame for the specific pan and tilt angles used during the camera calibration. The mast calibration described below is needed to compute the accurate camera model for any given pan and tilt angles.

C. Rover Mast Calibration

Our mast calibration method is an extension of the camera calibration previously described, viz. collect camera images and 3-D target positions at several different calibration target positions with different mast pan and tilt angles. The key idea is again to obtain 3-D metrology data relative to the same rover reference frame used in camera calibrations. This is critical to achieve accurate camera handoff 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 definition of the rover mast kinematics. In the Rocky8 rover, the Pancam/Navcam masthead is mounted on a vertical mast with a 2-DOF 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 rover reference frame by

$$T_{rover-to-camera} = T_{mast} * T_{masthead} * T_{camera} , \tag{6}$$

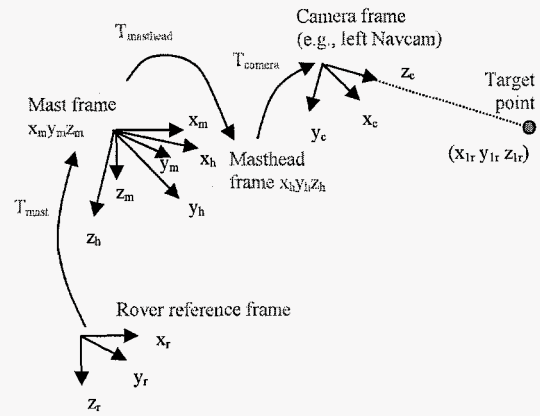

Fig. 3 Coordinate frame definitions for rover mast kinematics.

where T_{mast} is the mast frame relative to rover, $T_{masthead}$ is the masthead frame relative to mast, and T_{camera} is the camera frame relative to masthead. 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. The transform from the rover reference frame to the mast frame can be described by three translation and three rotation parameters:

$$T_{mast} = Trans(t_{xm}, t_{ym}, t_{zm}) * Rot(x, \theta_{xm}) * Rot(y, \theta_{ym}) * Rot(z, \theta_{zm}) , \tag{7}$$

where an ideal perfect straight-up mast will have zero rotation angels. Initially we added *pan_offset* and *tilt_offset* parameters to pan and tilt angles in representing $T_{masthead}$:

$$T_{masthead} = Rot(z, pan + pan_offset) * Rot(y, tilt + tilt_offset) . \tag{8}$$

Since both *pan_offset* and θ_{zm} are about the same rotation axis, it is clear that *pan_offset* was redundant and not needed. Further, when the non-linear least squares method was applied to determine seven parameters: t_{xm}, t_{ym}, t_{zm}, θ_{xm}, θ_{ym}, θ_{zm}, *and tilt_offset*, we found that *tilt_offset* is also a free variable that can be set to any value mathematically. Thus, in this paper we can safely assume that *pan_offset* = *tilt_offset* = 0, although the user can set them to any desirable values.

Input data to the mast calibration are 3-D points, their corresponding 2-D image points, pan and tilt angles, and camera models together with the pan and tilt angles used during the camera calibrations. Given the input data, nonlinear least squares determines the mast calibration parameters ψ^* that minimize the mean square error between the projected image points of the 3-D points and their corresponding 2-D image points.

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

where ψ is the six-parameter vector to solve

$$\psi = [t_{xm}\ t_{ym}\ t_{zm}\ \theta_{xm}\ \theta_{ym}\ \theta_{zm}] . \tag{10}$$

X_{3d} and X_{2d} are vectors of 3-D points with respect to the rover frame and their corresponding 2-D points. K_o is the camera model for each camera relative to the rover frame,

obtained by the camera calibration procedure described previously. All mast-mounted cameras must provide the pan and tilt angles pan_{Ko} and $tilt_{Ko}$ used during the camera calibration, since the camera calibration is done relative to the rover frame. The C function projects rover-frame-referenced 3-D points onto the image plane, which requires the camera model relative to the rover frame for given pan and tilt angles.

The mast calibration algorithm determines the six mast calibration parameters as well as the CAHVOR models relative to the masthead, which are fixed regardless of pan and tilt angles. For given pan and tilt angles together with the mast calibration parameters, the user and other application software can compute the mast-mounted camera models relative to the rover frame. Transformation of the CAHVOR camera model from one frame to another is achieved by the following: translate and rotate the **C** vector, only rotate **A**, **H**, **V**, and **O** vectors, and do not change the nonlinear parameter vector **R**. In the standard camera model [12], transformation is straightforward, since the camera model explicitly uses the extrinsic camera parameters representing the camera frame position and orientation.

We implemented the above mast calibration algorithm in MATLAB and tested it with the Rocky8 pan/tilt mast. Initially we used a total of 18 calibration target positions, but 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 -0.8 radians.

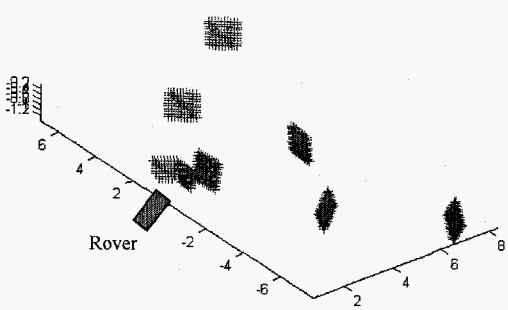

Fig. 4 Calibration target positions used for mast calibration.

TABLE IV
MAST CALIBRATION PARAMETERS IN METERS AND RADIANS

mast t_{xm}	*mast* t_{ym}	*mast* t_{zm}	*mast* θ_{xm}	*mast* θ_{ym}	*pan offset*
0.3443	-0.0043	-1.4572	-0.0088	-0.0181	0.0027

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

	Pancam residual error		Navcam residual error	
	x	y	x	y
6 parameters	0.854	1.102	1.129	1.002
4 parameters	2.871	22.294	2.784	7.934

Table IV lists the six 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 six-parameter mast calibration, the maximum residual rms error was 1.13 pixels. However, if the mast is assumed to be exactly vertical or perfectly parallel to 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 6 parameters to attain high accuracy.

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. A smaller search area improves the tracking reliability. For a given rover pose and target position estimates, the best accuracy in 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 in order 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 **A**) of the CAHVOR camera model in general does not pass through the center of the image. We found a non-iterative closed-form exact solution, which is presented here.

We first convert the target point from rover frame to mast frame, since T_{mast} is known by the mast calibration.

$$\begin{bmatrix} x_{1m} \\ y_{1m} \\ z_{1m} \\ 1 \end{bmatrix} = T_{mast}^{-1} \begin{bmatrix} x_{1r} \\ y_{1r} \\ z_{1r} \\ 1 \end{bmatrix}. \tag{11}$$

From (6) and (11),

$$\begin{bmatrix} x_{1m} \\ y_{1m} \\ z_{1m} \\ 1 \end{bmatrix} = T_{masthead} * T_{camera} \begin{bmatrix} x_{1c} \\ y_{1c} \\ z_{1c} \\ 1 \end{bmatrix}. \tag{12}$$

By denoting $s_1 = sin(pan)$, $c_1 = cos(pan)$, $s_2 = sin(tilt)$, $c_2 = cos(tilt)$ with $pan_offset = tilt_offset = 0$, (8) becomes

$$T_{masthead} = \begin{bmatrix} c_1 & -s_1 & 0 & 0 \\ s_1 & c_1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} c_2 & 0 & s_2 & 0 \\ 0 & 1 & 0 & 0 \\ -s_2 & 0 & c_2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{13}$$

In an ideal case when the camera frame is parallel to the masthead frame except that their axes are swapped, the camera frame is determined by first translating the masthead frame and then swapping the x, y, and z axis with the z, x, and y axis, respectively.

$$T_{camera} == \begin{bmatrix} 0 & 0 & 1 & t_{xc} \\ 1 & 0 & 0 & t_{yc} \\ 0 & 1 & 0 & t_{zc} \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{14}$$

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

$$x_{1c} = y_{1c} = 0. \tag{15}$$

From (11)-(15), one can solve for pan and tilt angles without much difficulty by both algebraic manipulation and geometric interpretation. In practical cases, however, the camera frame axes are not in general parallel to the masthead frame axes, and the camera optical axis in general does not pass through the center of the camera image.

When the optical axis of the camera model 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) lets the target point reappear at the center of the camera image. The CAHVOR camera model relative to the masthead defines its camera frame T_{camera} by

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

where **H** and **V** are not necessarily mutually orthogonal, while **H′**, **V′**, and **A** are mutually orthogonal unit vectors along the x, y, and z axis, respectively. In general, **H′** and **V′** can be extracted from the CAHVOR model.

$$\mathbf{H} = h_s\mathbf{H'} + h_c\mathbf{A}, \tag{17}$$

$$\mathbf{V} = h_s\mathbf{V'} + h_c\mathbf{A}, \tag{18}$$

Parameters h_s and v_s are horizontal and vertical focal lengths expressed in pixels. The optical axis vector **A** is perpendicular to the image plane, and passes though 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

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

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

In the standard (or Tsai) camera model, h_c, h_s, v_c and v_s are readily available in the 3×3 upper triangular matrix

$$K_i = \begin{bmatrix} h_s & \sigma & h_c \\ 0 & v_s & vc \\ 0 & 0 & 1 \end{bmatrix}. \tag{21}$$

When the offset amounts are computed by (19)-(20), the transform needed for offset correction is given by

$$T_{offset} = Rot([-\theta_v, \theta_h, 0]), \tag{22}$$

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. From (16) and (22),

$$T_{camera} * T_{offset} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_{xc} \\ r_{21} & r_{22} & r_{23} & t_{yc} \\ r_{31} & r_{32} & r_{33} & t_{zc} \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{23}$$

By inserting the offset correction transform of (22) and (23) into (12) and applying constraint (15), we have

$$T_{masthead}^{-1} \begin{bmatrix} x_{1m} \\ y_{1m} \\ z_{1m} \\ 1 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_{xc} \\ r_{21} & r_{22} & r_{23} & t_{yc} \\ r_{31} & r_{32} & r_{33} & t_{zc} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ z_{1c} \\ 1 \end{bmatrix}. \tag{24}$$

From (13) and (24), we can derive closed-form exact kinematic solutions for pan and tilt angles such that the target image appears at the center of the camera image. The actual derivation is somewhat lengthy and thus omitted in this paper, but in essence it involves solving a quadratic equation of the form

$$(d_{xy}\sin\beta)^2 + 2C(d_{xy}\sin\beta) + D = 0. \tag{25}$$

We first verified with MATLAB that the solutions were within sub-pixel accuracy. Thereafter, we performed actual mast camera pointing experiments (Table VI). After taking the Navcam stereo camera images, the user mouse-clicks the desired pointing target in the left image. The system then computes the 3-D position of the designated target and points the mast so that the target image appears at the center of the Navcam left image. After the first mast camera pointing, the system does one more pointing. This pointing test is repeated over 50 random target points for three different image sets. As shown in Table VI, the RMS pointing error was within 1.3 pixels for the first camera pointing and 0.9 pixels for subsequent pointing. The pointing accuracy is excellent, and a second pointing correction is completely unnecessary.

TABLE VI
CAMERA POINTING RMS PIXEL ERRORS FROM FISRT AND SECOND POINTING

Camera aiming distance	First Camera-Pointing rms pixel error		Second Camera-Pointing rms pixel error	
	Δx	Δy	Δx	Δy
10 m	0.591	1.219	0.527	0.778
6 m	0.743	1.243	0.596	0.844
2 m	1.260	0.925	0.672	0.921

V. CAMERA HANDOFF

As described in Section II.A, camera handoff is essential to achieve high accuracy in target approach. In an example baseline operation scenario, the target is initially tracked in Pancam images. On the way, it is switched to track in Navcam images, and finally switched to track in 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 MATLAB test code for the purely geometric method that uses stereo camera views. The computational procedure is 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.

As a way to measure the camera handoff accuracy, we used bricks with reflective-tape targets attached at their four corners. This method provides two benefits: 1) it enables accurate measurement of 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 handoff and at about 2 m away for Navcam-to-Hazcam handoff. 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 integer coordinate values, it inherently has ±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 the Pancams, we compute the target image positions in the Navcams by the purely geometric method. The cross marks overlaid on the Pancam image in Fig.5 are human-entered target positions, while those on the Navcam image are determined by the purely geometric method. The difference between the computed and human-entered target positions for the Navcam image is computed as the camera handoff error. The handoff error for each brick target for both left and right images is plotted in Fig. 6. The maximum error was 2.5 pixels, but there was a bias of about 1.5 pixels due to the discrepancy in reference frames between camera calibrations. Inaccuracy in human data entry contributes to slight exaggeration of the handoff error.

A typical test run of a Navcam-to-Hazcam handoff is shown in Fig. 7. The cross marks overlaid on the Navcam image are human-entered target image positions, while those on the Hazcam image are the corresponding target

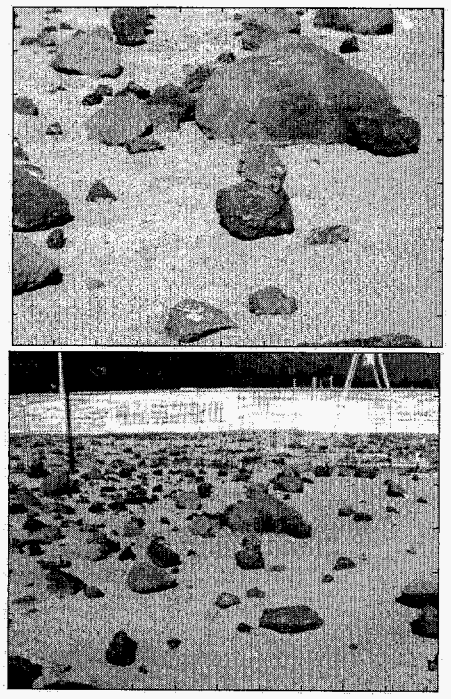

Fig. 5 Camera handoff from Pancam (top) to Navcam (bottom).

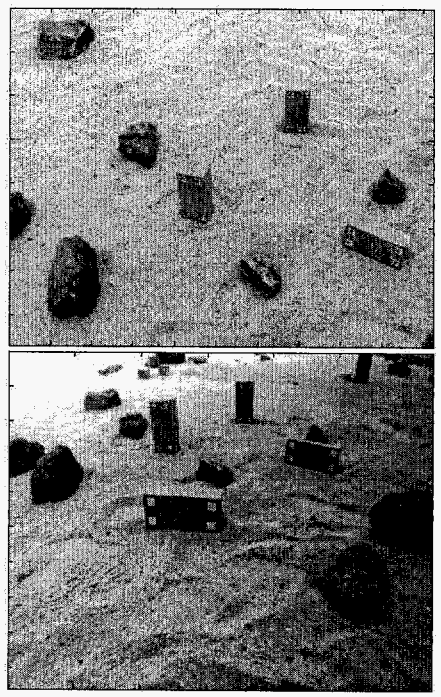

Fig. 7 Camera handoff from Navcam (top) to Hazcam (bottom).

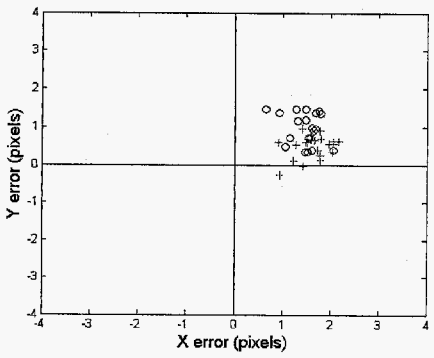

Fig. 6 Pancam-to-Navcam handoff error.

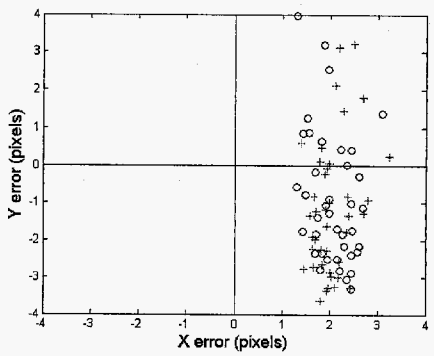

Fig. 8 Navcam-to-Hazcam handoff error.

positions determined by the purely geometric method. The handoff error for each brick target for both left and right images is plotted in Fig. 8. The maximum error was about 4 pixels, but there was a bias of about 2 pixels. Since the Hazcam in the current Rocky8 set-up is a lower resolution of 640×480, the equivalent maximum handoff error for a 1024×768 resolution is 4×1024/640 = 6.4 pixels.

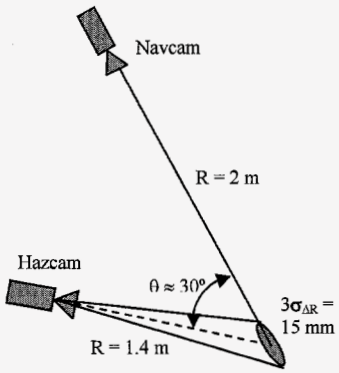

Fig. 9 Propagation of the Navcam stereo range error ellipsoid to Hazcam.

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, the Navcam-to-Hazcam handoff error in Fig. 8 shows a larger vertical spread. This vertical spread can be explained by observing that there is a significant difference in viewing directions between Navcam and Hazcam during the Navcam-to-Hazcam handoff. Fig. 9 illustrates how the major-axis of the Navcam stereo range error ellipsoid is projected onto the Hazcam image. The vertical projection length of the major-axis of the error ellipsoid onto the Hazcam image is approximately proportional to $\sin\theta$, where θ is the angle between Navcam and Hazcam optical axes (viewing directions). As the θ angle increases, more error is propagated to the Hazcam. In the current Rocky8 set-up, the angle change in the vertical direction is about 30° between Navcam to Hazcam, while it is nearly 0° between Pancam and Navcam. This is why we see the vertical error spread only for Navcam-to-Hazcam handoff.

In addition to the purely geometric handoff described above, we have experimented with a refinement step using normalized cross correlation. This is especially useful for Navcam-to-Hazcam handoff, since using an appearance based refinement method offsets some of the geometric uncertainty associated with mast calibration. However, we must compensate for the potentially large variation in target appearance due to viewpoint and field of view changes. Our approach is to project Hazcam stereo data into both the Hazcam and Navcam imagery to associate pixels between the two cameras. This allows a warped Navcam template in the neighborhood of the target to be mapped onto the Hazcam image for correlation and precise handoff. Our early tests have been promising and exhibit low handoff errors on the order of 1 pixel, provided the target is in an area of relatively high texture and away from occluding boundaries.

VI. Conclusion

We developed three technical elements that improve the accuracy of the visual target tracking approach system for use in single-sol approach and instrument placement in future Mars rover missions. The key elements developed are 1) rover mast calibration, 2) exact camera pointing, and 3) camera handoff. Algorithms were implemented in MATLAB, 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 Bajracharya, 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.

[12] R. Lenz and R. Tsai, "Techniques for Calibration of the Scale Factor and Image Center for High Accuracy 3-D Machine Vision Metrology," IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 10, no.5, pp. 713-720, 1988.

[13] Heikkilä, J, "Geometric Camera Calibration Using Circular Control Points", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, No. 10, pp. 1066-1077, Oct 2000.