Target Tracking, Approach, and Camera Handoff For Automated Instrument Placement

Max Bajracharya, Antonio Diaz-Calderon, Matthew Robinson, Mark Powell
Jet Propulsion Laboratory
4800 Oak Grove Dr.
Pasadena, CA 91109, USA
firstname.lastname@jpl.nasa.gov

Abstract—This paper describes the target designation, tracking, approach, and camera handoff technologies required to achieve accurate, single-command autonomous instrument placement for a planetary rover. It focuses on robust tracking integrated with obstacle avoidance during the approach phase, and image-based camera handoff to allow vision-based instrument placement. It also provides initial results from a complete system combining these technologies with rover base placement to maximize arm manipulability and image-based instrument placement.

TABLE OF CONTENTS

1. INTRODUCTION

In order to increase science return or decrease surface time, future Mars rover missions will need to be able to designate, safely approach, and accurately place an instrument on a target in a single sol (Mars day). This paper focuses on the approach, tracking, and handoff technologies required to achieve this capability and describes a complete end-to-end demonstration of the integrated system.

Because the communication window to a Mars rover is limited during the course of a day, autonomous operation of the rover is essential to increasing its productivity. Currently on the Mars Exploration Rover (MER) mission, the process of designating a target, approaching it, and placing an instrument on it takes a minimum of 3 communication cycles (3 sols) (one to approach it, one to refine the rover placement, and one to place the instrument). This paper describes the technology to allow a scientist to designate a target (generally a location on a rock) in a panorama with one command and then autonomously approach the target avoiding any hazards, refine the rover placement to maximize the manipulability of the arm, and place the instrument on the target.

All experiments described were carried out in JPL's Mars Yard (an outdoor terrain simulating the rock densities at the Viking 1 and 2 landing sites on Mars) on the Rocky 8 research rover. The Rocky 8 rover is a MER-like rover, with a six-wheel rocker-bogie suspension, a fixed pan/tilt mast with panorama and navigation stereo camera pairs, and front and rear body-mounted hazard stereo camera pairs. However, unlike the MER rovers, it has full six-wheel steering allowing it to execute crab maneuvers. The rover and it's environment can be seen in Figure 1.

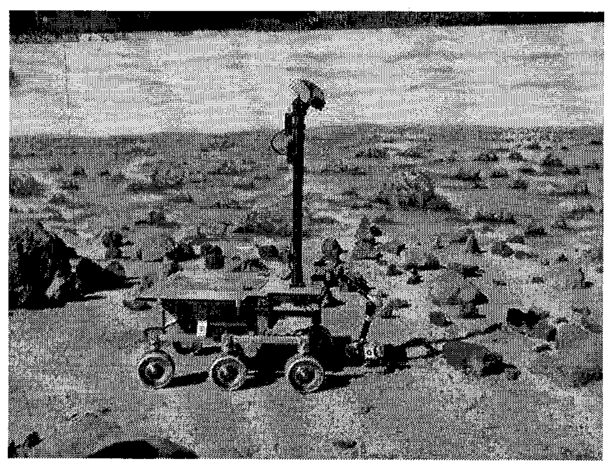

Figure 1. The Rocky 8 rover in the JPL Mars Yard

The experiment starts with the rover taking a panorama of the terrain with its mast cameras pointing out to ten meters on the ground. After taking the panorama, the images are downlinked to Maestro, the ground software system. Maestro is then used to designate a target in one of the images and generate a command to approach and place an instrument on that target. After receiving the command, the rover calculates an initial goal position such that the target is in the workspace of the arm, and uses the Morphin navigation algorithm (similar to the Gestlat algorithm being used on MER) to avoid obstacles and achieve the goal location. During approach, the rover also tracks the target in its mast cameras. When the

0-7803-8870-4/05/$20.00/© 2005 IEEE
IEEEAC paper # 1631

rover achieves its goal location, the target is projected from the mast camera to body cameras and the rover position is refined to maximize the arm manipulatibilty. During the refinement motions, the target is tracked in the body cameras. When the final position is achieved, the arm is deployed and the instrument placed on the target using feedback from the body cameras.

The system that was developed is an integration and improvement of many previously developed technologies. Target designation makes use of Maestro, an improved version of WITS/SAP [12]; mast pointing uses the kinematics derived in [3]; tracking uses concepts developed from previous work at JPL [4] and at NASA Ames Research Center [14]; navigation uses an implementation by CMU [16]; camera handoff uses an idea from [3]; and rover base and instrument placement is described in a companion paper [15]. An integrated system with similar capabilities using several different technologies has also been developed at NASA Ames Research Center and is described in [10].

2. Target Designation

The scenario begins with the designation of the target in the ground systems software. Maestro is a science analysis and activity specification tool used for both research rovers and mission spacecraft. It provides immersive visualization to quickly understand the terrain in the vicinity of a rover and also to specify activities for the science instruments, manipulators, and mobility systems to execute. It is used to drive the Rocky 7, Rocky 8, and FIDO technology rovers and has supported flight missions including Mars Polar Lander and Mars Exploration Rover (for which it was named the Science Activity Planner). We use Maestro to specify a single cycle instrument placement command for execution on the Rocky 8 rover. After the rover has taken a panorama and downlinks the images, an operator loads the images in Maestro. After viewing the images to understand the local terrain, the operator designates a location on the soil or on a rock to place the instrument by creating a target. For our work, the target is specified as a particular pixel (i, j) coordinate in an image where we want to place an instrument. Finally, the operator creates an activity to place the instrument and uplinks the activity to the rover.

3. Approach

After receiving the command, the rover uses the designated target and calculates the 3D location of the target and computes an initial rover base placement to achieve (this algorithm is described in [15]). Using the initial base placement as a goal, the rover then uses the Morphin navigation algorithm to approach the target. During navigation, the rover also tracks the target by pointing the mast at the target, predicting the location of the target in the image, and then template matching.

Navigation

The Morphin navigation algorithm [17] analyzes range data from the front and rear body cameras to avoid obstacles and achieve the specified goal location. Morphin bins the range data into a gridded map and uses statistical metrics of slope, roughness, and step height to determine traversability. It then uses a local and global cost function to analyze potential arc motions, integrating the cost along the arc and choosing the best arc and distance along that arc. In our system, we use arc lengths between 0.5 and 1.5 meters and parameters specific to the Rocky 8 rover mobility systems (e.g. traversable step sizes of 20cm).

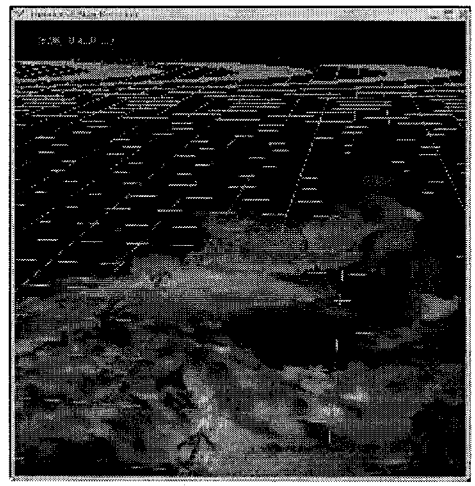

Figure 2. The Morphin navigation map and rover poses

Tracking

As the rover approaches the target, visually tracking the target is essential in order to eventually place an instrument on the target accurately. Although vision based pose (position and orientation) estimation can be quite good, on a 10m straight line drive even a 1% error in pose estimation would result in a 10cm error in instrument placement (a yaw error accumulated over the traverse could potentially contribute to significantly more). Therefore, in order to place an instrument to several centimeters accuracy, tracking the target to within several pixels in an image is critical.

Because of flight hardware constraints, Mars rovers generally do not have the ability to acquire images continuously and consequently use stop and go motions (of about half a meter). As a result, standard tracking techniques must be modified to work without a small motion assumption. In addition, our tracking algorithm uses a combination of methods to improve robustness and prevent drift.

Using the rover's pose estimation and the triangulated 3D location of the target, the cameras on the mast's pan/tilt head can be pointed at the target, reducing the motion of the target in the image. This requires the kinematic solution to the problem of pointing a camera with a fixed transformation (translation and rotation) from its pan/tilt axes to a 3D point in the world. The goal is to find the pan/tilt angles that project the 3D point to the center of the image, taking into account the camera model parameters. A complete description to the solution to the inverse kinematics problem of exact camera pointing is given in [3]. Briefly, knowing the transformation of an image coordinate $[x_i y_i z_i]$ to a world point $[X_p Y_p Z_p]$, we can then find the pan and tilt angles in the masthead to camera transformation ($T_{camera}^{masthead}$) such that a known world point projects to an image location of $x_i = 0$, $y_i = 0$.

$$\begin{bmatrix} X_p \\ Y_p \\ Z_p \end{bmatrix} = T_{masthead}^{rover} * T_{camera}^{masthead} * T_{image}^{camera} \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} \quad (1)$$

Setting $x_i = 0$ and $y_i = 0$, and knowing the other transformations, we can then solve for the rotation angles (resulting in a quadratic solution) (see [3] for details).

While the inverse kinematics for centering the target in the image is exact, inaccuracies in both rover pose and mast kinematics can contribute to a significant error in pointing. If pointing was always within a few pixels, local gradient descent search methods would be sufficient to match and track the target. However, because of point inaccuracies, an initial global matching method must be used.

To overcome the pointing error, a fast normalized cross correlation search [13] at a low pyramid level is used between sequential frames. Normalized cross correlation search simply seeks the location in some window that maximizes the correlation:

$$\sum_{x,y} [I(x,y) - \bar{I}_{x',y'}][I'(x - x', y - y') - \bar{I}'] \quad (2)$$

Because this matching process only takes into account translation, it must be done often enough to avoid a significant scale or perspective change in the target. In addition, because the template must be continuously updated between frames, if used solely, would tend to drift. Because the accuracy of this initial matcher is not critical, a larger template can be used. While a larger template tends to include depth discontinuities, because it also includes the context of the rock rather than simply the texture, it is less prone to false matches.

To address the accuracy and drift problems of the normalized cross correlation matcher, it is augmented with a more accurate local matcher which takes into account an affine transform (translation, scale, and skew, approximating a perspective transform) of the original target template ([7], [5]).

The goal of matching is to find the translation/transformation of a window from one image in another image (with a small motion assumption between the images). We accomplish this by solving for the translation/transformation parameters ($\delta\mathbf{D}$, $\delta\mathbf{t}$) that minimizes the squared intensity error between the two image windows I and I' at pixel i:

$$E(\delta\mathbf{D}, \delta\mathbf{t}) = \sum_i [I(\mathbf{x}_i) - I'((\mathbf{I} + \delta\mathbf{D})\mathbf{x}'_i + \delta\mathbf{t})]^2 \quad (3)$$

where:

$$\mathbf{x}_i = (x_i, y_i) \quad (4)$$

and:

$$\mathbf{x}'_i = (x'_i, y'_i) = (\mathbf{I} + \mathbf{D})\mathbf{x}_i + \mathbf{t} \quad (5)$$

In order to minimize this equation, we must first take its first order Taylor expansion; combining the unknowns $\delta\mathbf{D}$ and $\delta\mathbf{t}$ into $\mathbf{v}$, we have:

$$E(\mathbf{v}) \approx \sum_i [\mathbf{g}_i^T \mathbf{J}_i^T \mathbf{v} + e_i]^2 \quad (6)$$

where:

$$\mathbf{g}_i = \nabla I(\mathbf{x}') = \begin{bmatrix} I_x \\ I_y \end{bmatrix} \quad (7)$$

$$\mathbf{J}_i = \frac{\partial \mathbf{x}'}{\partial \mathbf{v}} \quad (8)$$

$$e_i = I(\mathbf{x}_i) - I'(\mathbf{x}'_i) \quad (9)$$

We can now solve the equation using a least squares method. Taking the partial derivative and setting equal to zero:

$$\frac{\partial E(\mathbf{v})}{\partial \mathbf{v}} = 2 \sum_i (\mathbf{g}_i^T \mathbf{J}_i^T \mathbf{v} + e_i)(\mathbf{J}_i \mathbf{g}_i) = 0 \quad (10)$$

We can then solve for $\mathbf{v}$:

$$\mathbf{v} = -\mathbf{G}^{-1}\mathbf{b} \quad (11)$$

where:

$$\mathbf{G} = \sum_i \mathbf{J}_i \mathbf{g}_i \mathbf{g}_i^T \mathbf{J}_i^T \quad (12)$$

and

$$\mathbf{b} = \sum_i e_i \mathbf{J}_i \mathbf{g}_i \quad (13)$$

The implementation actually leaves the original feature window the same and translates/transforms the new image so that the gradient and Jacobian only need to be computed once, similar to the implementation described in [6].

1. For each feature window, at each pyramid level
   (a) Calculate gradient $\mathbf{g}$
   (b) Calculate Jacobian $\mathbf{J}$
   (c) Calculate $\mathbf{G}^{-1}$
2. For each new image, at each pyramid level, starting at the lowest resolution, iterate the following steps (where the next pyramid level iteration is initialized with the previously calculated $\mathbf{v}' = \frac{\mathbf{v}}{2^{level}}$):
   (a) Calculate $\mathbf{b}$ using $\mathbf{v}'$
   (b) Solve for $\mathbf{v}$
   (c) Update $\mathbf{v}' = \mathbf{v}' + \mathbf{v}$

For translation and an affine transform, we have:

$$\mathbf{D} = \begin{bmatrix} d_0 & d_1 \\ d_2 & d_3 \end{bmatrix} \quad (14)$$

and so the Jacobian is:

$$\mathbf{J} = \begin{bmatrix} x & 0 \\ y & 0 \\ 0 & x \\ 0 & y \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \quad (15)$$

Because only the template parameters are updated at each frame, and the template itself is only updated periodically, this tracker drifts less than a correlation approach.

Because only one feature is being tracked, some additional computational time can be used the track the feature more robustly. The matcher is run using multiple template sizes (to retain high frequency information at larger sizes), as well as multiple pyramid levels to speed up the matching. At each size and level, the change in parameters is constrained to be within the theoretical limit of the matcher (at the highest pyramid level, with a two pixel difference as the derivative, this would be one pixel). If the drift is higher than anticipated, the previous result is retained. In addition, the 3D location of the target is calculated and compared to the previous matcher to verify that it is in the same location. The first and last images of a target being tracked in the body cameras is shown in figures 5 and 6.

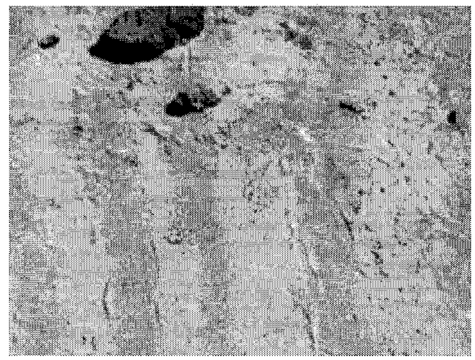

Figure 3. The target location in the mast cameras

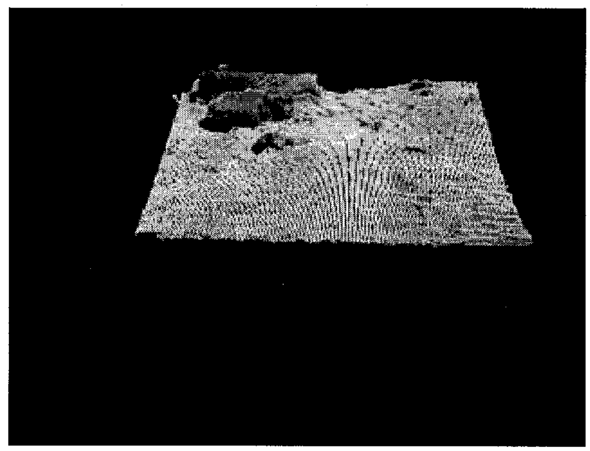

Figure 4. The projection of the mast cameras to the body cameras

4. HANDOFF

After achieving the initial rover base placement and tracking the target up to this location, transferring the target from the mast cameras to the body cameras is necessary to continue tracking the target, refine the base placement, and eventually place the instrument using an image based algorithm. Although projecting the target from the mast cameras to the hazard cameras can theoretically provide accuracy within several pixels, because of the sensitivity to camera and kinematic calibration, augmenting the process with a more accurate image based template matching method was necessary. This method uses the dense stereo produced from the mast cameras to create a projected body camera image (see [3] for more details). The original mast image can be seen in figure 3, and the projection to a body camera can be seen in figure 4 (the black pixels indicate areas where no data is available, either due to the projection or missing stereo data; the missing data is ignored when correlating the projection to the real image); the real body camera image is shown in figure 5. We then match a template of the target in the projected image (figure 4) to the

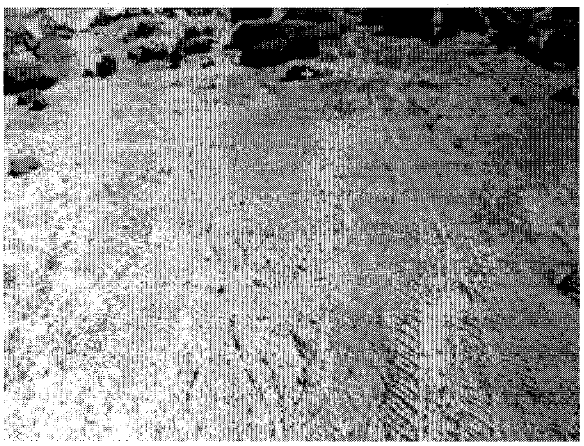

Figure 5. The initial target location in the body cameras

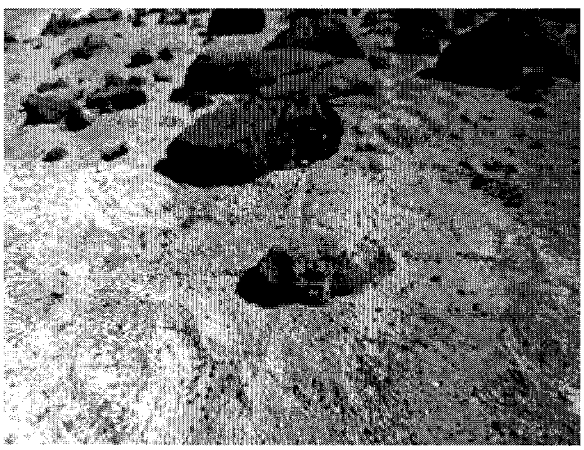

Figure 6. The final tracked target location in the body cameras

real body camera image (figure 5). This accurately localizes the target in the body camera by using both the features texture as well as its context. The template size can be chosen as large as possible given the projection and computational constraints because the images appear to be from the same point of view.

5. ROVER BASE AND INSTRUMENT PLACEMENT

Once the target is localized in the body cameras, the base placement of the rover is iteratively refined, recomputing the targets 3D location and moving the rover to maximize arm manipulability on each iteration.

Tracking during the base placement motion uses the same algorithm as during approach, but since the body cameras cannot be pointed, it uses the estimated rover motion to predict the location of the target in the image and seeds the tracker with this location. The result of tracking in the body camera is shown in figure 6.

Once the rover is placed such that the target is in the workspace of the arm, a vision based technique is used to accurately place the instrument on target. Rover base placement and instrument placement are described in the companion paper [15].

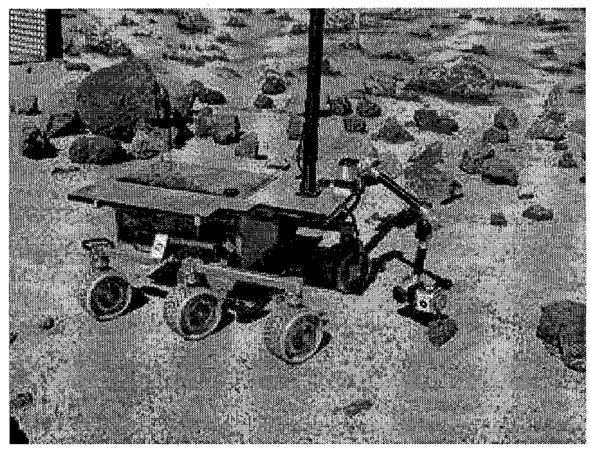

Figure 7. The rover position after iterative base placement, with the arm extended

Figure 8. The placement of the instrument on the tracked target

6. INITIAL RESULTS AND VALIDATION

Initial experiments of the integrated end-to-end system in the JPL Mars Yard have shown that an accuracy of several centimeters can be achieved for a target approximately 10 meters away. A task is currently validating and determining the accuracy and repeatability of the entire system. The potential sources of error (approach tracking, handoff, base placement tracking, and instrument placement) have all been identified but not all rigorously tested. Using the mast cameras, tracking between frames is generally accurate to less than a pixel, and less than several pixels over the entire 10m traverse. The 2D tracking algorithm has gone through a validation process ([1]) and shown to be as accurate as 4cm (3σ), but has been augmented in this work to be more robust to failure.

Because tracking and an iterative final base placement approach is used, overall pose estimation is not critical. Relative pose estimation between tracking frames simply needs to be good enough so that the cameras can be pointed such that the target is within several hundred pixels from the center of the image (as this is the search window for the normalized cross correlation algorithm). In addition, the the navigation algorithm only needs to achieve the initial goal to within a 1m radius as the final placement is refined by the base placement procedure. Finally, using image based instrument placement (HIPS) reduces the placement accuracy error from several centimeters to less than one centimeter ([2]).

We believe that the largest contribution to overall instrument position error in this system is currently the handoff technique being used. Because of the change in view point, matching the reprojected mast cameras image to the body camera image is only accurate to a few pixels. A potentially more accurate method would be to use the 3D information from the two cameras directly, registering the point clouds being generated ([11]).

7. CONCLUSION

This paper describes a system capable of accurately placing an instrument on a target designated from ten meters away. Critical to achieving this accuracy is the ability to track the target in the pointable mast cameras, the capability to handoff the target from the mast cameras to the body cameras, and the ability to track the target in the non-pointable body cameras. Additionally, the technologies of navigation, kinematic mast pointing, and rover base and instrument placement are necessary to integrate the complete system.

The ability to autonomously place an instrument with a single command will significantly increase the productivity of future rover-based Mars missions. Even achieving a few centimeters accuracy can improve the science return by sampling or observing the target to provide additional information for the next placement. When minimizing time on the surface becomes a mission priority (as expected for the 2013 Mars Sample Return mission), single cycle instrument placement will be a necessary technology to enable the mission.

8. ACKNOWLEDGEMENTS

The research described in this paper was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

REFERENCES

[1] W. Kim, R. Steinke, R. Steele, "2-D Target Tracking Technology Validation Report", JPL Technical Report D-28523, Apr. 2004.

[2] P.Backes, "Base Placement..." *IEEE Aerospace Conference*, Big Sky, Montana, 2005.

[3] W.S.Kim, A.I.Ansar, R.D.Steele, "Rover Mast Calibration, Exact Camera Pointing, and Camera Handoff for Visual Target Tracking," *IEEE International Conference on Robotics and Automation (ICRA)*, May 2005.

[4] 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.

[5] B.D. Lucas, T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision", *IJCAI*, 674-679, 1981.

[6] J. Bouguet, "Pyramidal Implementation of the Lucas Kanade Feature Tracker: Description of the Algorithm," Intel Corporation, Microprocessor Research Labs, 2000.

[7] J. Shi, C. Tomasi, "Good Features to Track", *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, June 1994.

[8] I. Nesnas, M. Maimone, H. Das, "Autonomous Vision-Based Manipulation from a Rover Platform," *IEEE Symp. on Computational Intelligence in Robotics and Automation*, 1999.

[9] T. Huntsberger, H. Aghazarian, Y. Cheng, E.T. Baumgartner, E. Tunstel, C. Leger, A. Trebi-Ollennu, P.S. Schenker, "Rover Autonomy for Long Range Navigation and Science Data Acquisition on Planetary Surfaces," *IEEE Int. Conf. on Robotics and Automation*, Washington, D.C., May 2002.

[10] L. Pederson, M. Bualat, D.E. Smith, R. Washington, "Integrated Demonstration of Instrument Placement, Robust Execution and Contingent Planning," *International Symposium on Artificial Intelligence and Robotics in Space (i-SAIRAS)*, Nara, Japan, 2003.

[11] M. Deans, C. Kunz, R. Sargent, L. Pederson, "Terrain Model Registration for Single Cycle Instrument Placement," *International Symposium on Artificial Intelligence and Robotics in Space (i-SAIRAS)*, Nara, Japan, 2003.

[12] P. Backes, J. Norris, M. Powell, M. Vona, R. Steinke, J. Wick. "The Science Activity Planner for the Mars Exploration Rover Mission: FIDO Field Test Results," *IEEE Aerospace Conference*, Big Sky, Montana, 2003.

[13] J. P. Lewis, "Fast normalized cross-correlation," *Vision Interface*, 1995.

[14] D. Wettergreen, H. Thomas, M. Bualat, "Initial results from vision-based control of the Ames Marsokhod

rover", *IEEE Int'l Conf. on Intelligent Robots and Systems*, France, 1997.

[15] P. Backes, A. Diaz-Calderon, M. Robinson, M. Bajracharya, D. Helmick, "Automated Rover Positioning and Instrument Placement," *IEEE Aerospace Conference*, Big Sky, Montana, 2005.

[16] C. Urmson, R. Simmons, I. Nesnas, "A Generic Framework for Robotic Navigation," *IEEE Aerospace Conference*, Big Sky, Montana, 2003.

[17] R. Simmons et al., "Experience with Rover Navigation for Lunar-Like Terrains," *Proc. Conference on Intelligent Robots and Systems*, Pittsburgh PA, August 1995.

Max Bajracharya *Max Bajracharya is a member of the Telerobotics Research and Applications Group at the Jet Propulsion Laboratory, Pasadena, CA. His research includes vision-based tracking, navigation, and pose estimation for rovers and he is currently a lead systems and software engineer for the CLARAty software project and the Rocky 8 research rover. He received his Bachelors and Masters degrees in computer science and electrical engineering from MIT in 2001.*

Antonio Diaz-Calderon *Antonio Diaz-Calderon is a member of Information Systems and Computer Science Staff with the Telerobotics Research and Applications Group at the Jet Propulsion Laboratory, Pasadena, CA. His research work at JPL involves manipulation kinematics, optimal estimation and motion control. He received the BSEE in 1987 and the M.S. in Computer Science in 1990 from Monterrey Institute of Technology (Mexico City, Mexico), and M.S. in 1993 in Engineering and Ph.D. in 2000 in Electrical and Computer Engineering from Carnegie Mellon University. He held a Post-Doctoral Fellowship appointment with The Robotics Institute at Carnegie Mellon University (2000-2001), and a faculty appointment as Commercialization Specialist with The Robotics Institute (2001-2003). While at the Robotics Institute Dr. Diaz-Calderon developed an estimation-based tipover sensing system to enhance dynamic stability of heavy articulated machines, robust position estimation algorithms for the PerceptOR program and vehicle dynamic models used for mobility analyses on prototype vehicles for Future Combat Systems' Unmanned Ground Combat Systems (UGCV), and Gladiator. His research interests include: sensor fusion and state estimation, and rover navigation and mobility.*

Mark Powell *Mark W. Powell has been a member of the technical staff in the Mobility Systems Concept Development section at the Jet Propulsion Laboratory, Pasadena, CA, since 2001. He received his B.S.C.S. in 1992, M.S.C.S in 1997, and Ph.D. in Computer Science and Engineering in 2000 from the University of South Florida, Tampa. His dissertation work was in the area of advanced illumination modeling, color and range image processing applied to robotics and medical imaging. At JPL his area of focus is science data visualization and science planning for telerobotics.*

Matthew Robinson *Matthew L. Robinson received the B.S. in 1996 and the Ph.D. in 2001, both in mechanical engineering from the University of Notre Dame. Currently, he is a member of the Engineering Staff in the Mechanical and Robotic Technologies Group at the Jet Propulsion Laboratory in Pasadena, CA. At JPL he has both flight mission and research experience. Currently, he contributes to the 2007 Phoenix mission (a lander equipped with a back-hoe style robotic arm and various scientific instruments destined for the Martian pole) as the flight software lead for the robotic arm. His research efforts have included the development of calibration-free, stereo-based manipulation techniques (HIPS) and control systems for holonomic and hybrid nonholonomic/holonomic. His research interests include machine vision, simulation, manipulation, mechatronic systems and controls.*

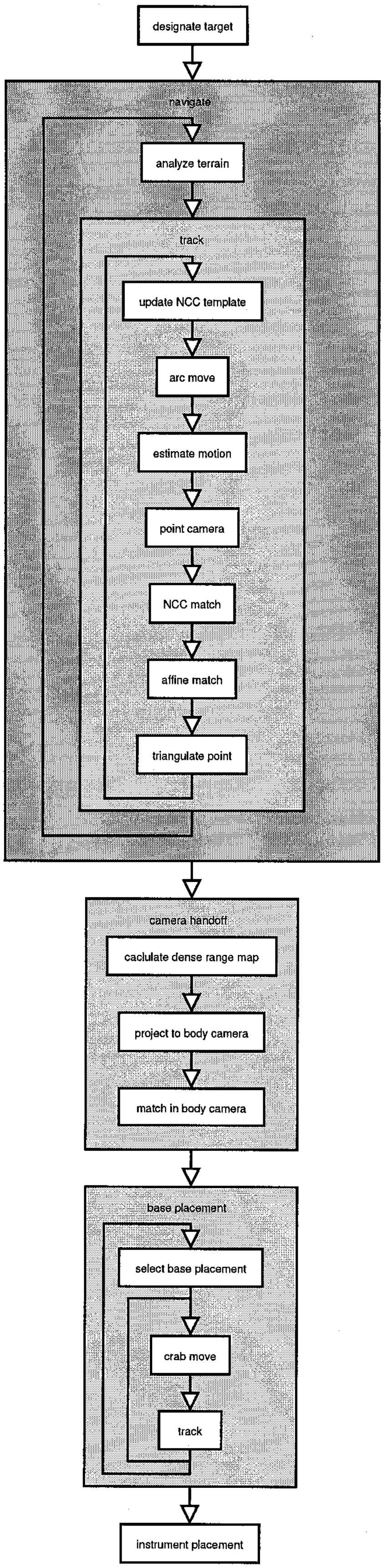

Figure 9. Algorithm Flowchart