To appear in International Symposium on Artificial Intelligence, Robotics and Automation in Space (i-SAIRAS), June 1999, Noordwijk, Netherlands

Autonomous Rock Tracking and Acquisition from a Mars Rover

Mark W. Maimone, Issa A. Nesnas, Hari Das
Jet Propulsion Laboratory
Pasadena, CA 91109
Email: mark.maimone@jpl.nasa.gov
Phone: +1 (818) 354 - 0592 Fax: +1 (818) 393 - 4085
http://robotics.jpl.nasa.gov/tasks/pdm/papers/isairas99/

Abstract

Future Mars exploration missions will perform two types of experiments: science instrument placement for close-up measurement, and sample acquisition for return to Earth. In this paper we describe algorithms we developed for these tasks, and demonstrate them in field experiments using a self-contained Mars Rover prototype, the *Rocky 7* rover. Our algorithms perform visual servoing on an elevation map instead of image features, because the latter are subject to abrupt scale changes during the approach. This allows us to compensate for the poor odometry that results from motion on loose terrain.

We demonstrate the successful grasp of a 5 cm long rock over 1m away using 103-degree field-of-view stereo cameras, and placement of a flexible mast on a rock outcropping over 5m away using 43 degree FOV stereo cameras.

1 Introduction

NASA is engaged in a series of missions designed to study the planet Mars. The current schedule calls for 5 pairs of orbiter/lander probes to be launched approximately every two years, starting with the Mars Pathfinder mission of 1997. The 2003 and 2005 missions, in particular, call for a rover with the ability to traverse more than 1 kilometer away from its landing site, acquiring samples along the way.

Autonomous robotic operations can greatly increase the science return of such planetary missions. As these operations become more adaptive, the burden of planning a sequence of motions is moved from the human operator to the onboard control system, allowing a greater number of targeted experiments to be achieved. In this paper we describe algorithms that allow a rover to autonomously approach and collect (or analyze) a sample at a human-specified target location.

Figure 1: The *Rocky 7* rover

Our approach combines vision processing with vehicle and arm control. The target is identified in an image by a human operator, and its 3D location is computed onboard using stereo vision. A curved path toward the target point is planned, and executed in small steps. The shape of the terrain immediately around the target is used to reacquire the target at each step; we servo on the elevation map instead of image features, because the latter are subject to abrupt scale changes during the approach. This allows us to compensate for the poor odometry that results from motion on loose terrain, by visually reacquiring the target at each step. Vehicle motion stops when the target appears within the workspace of the arm that will be used to grasp or study it.

In the sections that follow, we survey related work that uses visual servoing to guide end-effector motion, describe the general algorithm, and detail the experimental results from field tests performed on the *Rocky 7* Mars Rover prototype (see Figure 1).

1.	Acquire stereo image pair with body navigation cameras
2.	Send the left image over wireless network to host
3.	Scientist/Operator selects target rock on left image
4.	Target location and intensity threshold sent to rover
.	*All subsequent processing occurs onboard*
5.	Identify 3-D location of rock based on calibrated camera models and onboard stereo image processing
6.	Compute single-arc rover trajectory to target
7.	Drive rover toward target
8.	Periodically (every 10 cm) poll the target tracking software to update target location using new stereo pair and current odometry
9.	Redirect rover toward the new target location using new single-arc trajectory, and repeat until target is within 1 cm of goal position.
10.	Deploy sampling arm and pick up rock.

Table 1: Algorithm for small-rock acquisition

2 Related Work

First described in [WSN85], visual servoing strategies incorporate vision sensing with the actuation of motors in a robotic system. Often simple image-processing filters are used to locate a target of interest, and knowledge of the camera system geometry and manipulator kinematics are used to control motor current. This technique has been applied successfully to the active placement of a manipulator at high frame rates (e.g., in [HGT95], [PK93], [Nis90], and [THM+96]). In this application the distance of the target from the camera system usually remains the same, so the relative size of the object will remain constant throughout the servoing process.

In our case the entire robot, not just a manipulator, is being directed toward a goal point. Visual servoing for vehicle motion should be a useful tool, because the uncertainties introduced by motion over unknown terrain could potentially be eliminated by the visual tracking. However, as the vehicle approaches the target, the target's image size grows dramatically between updates, and a correlation search on the intensity image tends to fail. Therefore approaches such as [WTB97] work well at long distances, but are less reliable at the final approach to the object.

3 Approach

The general problem we attempted to solve is the identification and collection of an interesting rock sample, in a control architecture that meets the constraints of interplanetary operation. This latter requirement is summarized as follows: there will be a high latency in communication between the operator and rover (from 4 to 21 minutes one-way), and the number of messages sent must be minimized. For example, during Mars Pathfinder operations in 1997, logistical constraints on the Deep Space Network dictated that only two 5-minute communications windows were available each day.

This general problem can be broken down into a series of steps: Target Selection, Rover Motion toward the Target, Target Visual Reacquisition (these two steps might repeat a number of times), and Target Grasping. The first of these steps, Target Selection, is an extremely difficult task to automate, because it would require the rover to determine which samples are scientifically interesting. We felt this was a task best left to scientists, and therefore designed our system to require a single round-trip transmission to allow a human scientist to perform it. We felt that the remaining steps could be made sufficiently robust to be implemented entirely onboard the rover.

A summary of our algorithm for sample collection can be found in Table 1. The following subsections describe each component of the algorithm in detail, and refer back to the numbered steps in Table 1.

3.1 Target Selection

Target Selection is the first step of our sample acquisition process (steps 1-4 in Table 1). We assume the rover is already deployed in the area of interest, and has taken a stereo pair of images of the terrain in front of it. We transmit the left image from this stereo pair over the wireless network to a human operator who inspects the image, locates an interesting sample (a surface rock small enough to be grasped by the robot arm), selects it with the mouse, and transmits its image location back to the rover. Figure 2 illustrates a sample target selection. This step requires one round-trip communication between the rover and operator.

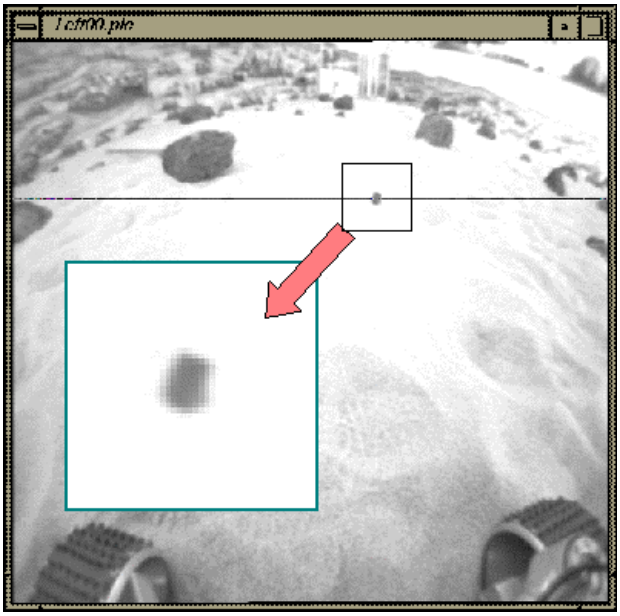

Figure 2: Sample target selection in Java GUI display. The selected target is shown zoomed in.

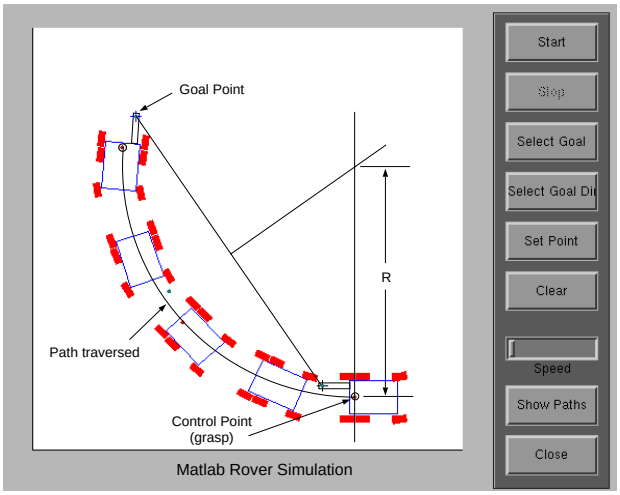

Figure 3: Single arc trajectory generation

We found it necessary in later processing to segment out the rock from its background using brightness-based intensity thresholding. So in addition to the image coordinates of the target rock, the operator communicates a brightness threshold and range to the rover (e.g., "pixels with 8bit intensity darker/lighter than 145 should be considered rocks").

3.2 Rover Motion toward the Target

Next the rover performs computations and moves toward its target (steps 5-7 and 9 in Table 1). Once the rover receives the goal point in image coordinates, it uses stereo image processing and a geometric camera model to compute the (X,Y,Z) location of the target in the rover reference frame. Details of the JPL Stereo Vision algorithm can be found in [XM97]. Note that the goal location is stored in the 3-D rover reference frame, not a 2-D image frame.

Having computed a location in world coordinates, a single arc is computed that should bring the rover close enough to the target that it appears within the workspace of the arm (see Figure 3). Our experimental arm had only 2 degrees of freedom, so it was important that the rover be positioned correctly to within a small tolerance, i.e., about 30% of the size of the 2 DOF gripper.

The rover is then commanded to move a short distance along the arc (10 cm or the remaining distance to goal, whichever is smaller), and its position is reevaluated in the next step.

3.3 Target Visual Reacquisition

Having made partial progress toward the goal, the rover stops to evaluate its current position (step 8 in Table 1). This update is initialized by subtracting the motion just taken from the target location in the rover frame. The motion just taken is estimated by computing vehicle odometry from wheel rotations. This is a very noisy estimate, because nothing is known about the surface on which the rover is moving; it could consist of pebbles, sand, sticky tar, or solid rock.

A starting point in a fresh stereo image pair is computed from this new estimated location, and a small window around that point is searched in an attempt to locate the target. However, instead of searching the raw intensity image we automatically compute a range image from the stereo image pair, and search the resulting elevation map for the *shape* of the target, rather than its visual appearance. In particular, we assume that any target rock will be resting higher on the ground than its nearby surroundings, and lock in on the local elevation maximum as the new, refined 3D target point. We may not always achieve a completely dense elevation map from the range data, so before searching for the local maximum we linearly interpolate any data missing from the range image. Given this dense, interpolated elevation map, we start at the best estimate of the target location and "climb" to higher elevations until we reach a local maximum.

Unfortunately, early experiments showed that on

a sandy surface, the error in the odometry estimate was sufficient to cause this method to lose the target. That is, the search window was centered too far away from the target rock for a simple gradient-ascent climb to recover it, even after relatively small motions. A general solution to this problem would be to incorporate more effective position and pose sensing and estimation into the rover. We anticipate that the work described in [Bal99] will provide such estimates and will be incorporated onboard the *Rocky 7* rover soon, but it was not available during the timeframe of our project.

Instead, we took advantage of the fact that our targets were visually distinct from the background sand, and used an intensity filter to focus attention in the elevation map. Given the search window centered at the (noisy) estimated target location, pixels in the image window are classified in one pass as either BACKGROUND or ROCK according to the threshold value set by the operator. The ROCK pixel nearest the center of the search window is then treated as part of the target, and the enclosing blob of ROCK pixels are relabeled TARGET pixels. Finally, the centroid of all TARGET pixels is computed, and its range value (perhaps an interpolated value) is used as the starting point for the climb to the local elevation maximum. Using the centroid preserves the scale-invariance of our method. In fact, any pixel classification technique can be used instead of brightness: on a flight mission one might use spectral filters to distinguish rocks from non-rocks, as in [PAW+98].

If no range data are available, then no refinement is done, and the vehicle odometry is assumed to be correct.

The new target location is fed back into the Rover Motion toward Target step, and vehicle motion continues until the target is found to be within the workspace of the arm.

3.4 Target Grasping

Finally, having determined that the target lies within the workspace of the arm, the arm is deployed and the target grasp is attempted (step 10 in Table 1). We use the difference between the actual and commanded trajectories from the motor encoders to tell when the arm makes contact with the target or ground, then close the gripper on the target. Instead of lifting off right away, we raise the arm a small amount and continue to close the gripper until it stops, several times more. This redundancy helps ensure that the gripper has a good hold on the target.

4 Experimental Results

As testbed for these algorithms, we used the *Rocky 7* Mars Rover prototype [Vol99] (see Figure 1). *Rocky 7* is a 6-wheeled vehicle with rocker-bogey suspension and one set of steerable wheels. Batteries and solar cells provide about 50 Watts of power. A small 2 DOF arm with 2 DOF gripper mounted on one side of the vehicle is used for digging and grasping rock samples, and an extendible 3 DOF mast provides stereo image views from as high as 1.5 meters above the ground. For terrestrial work, communication is via a 1 Mbit/sec wireless ethernet bridge or a 10 Mbit/sec coax hard line. Onboard processing consists of a 60 Mhz 68060 CPU running the VxWorks 5.3 operating system in 16 megabytes of RAM. Vision sensors include three pairs of stereo cameras: one body-mounted pair faces the arm, another body-mounted pair is on the other side of the vehicle, and the third pair is mounted near the end-effector on the extendible mast. All cameras are 480x512 CCD board cameras (but currently only half-resolution images are used), and the body-mounted cameras have an effective FOV of 103 degrees, while the mast cameras have an effective FOV of 43 degrees. The body-mounted cameras are approximately 30 cm above the ground, point downward at an angle of approximately 45 degrees, and are used primarily for detection of nearby obstacles. During these experiments the vehicle moved approximately 5 cm/sec and paused briefly during the image acquisition and path generation steps.

We performed several experiments in JPL's Mars Yard[1], and successfully demonstrated the autonomous acquisition of small rocks (3-5 cm) located over 1 meter in front of the rover. Figure 4 shows a sample tracking sequence, with the target indicated in each frame by a dark square. Execution of the entire sequence (Target Selection, $8-10$ iterations of Target Reacquisition, and successful Target Grasping) typically completed within one minute when the target was just over 1 meter away.

Many experiments were run, and 14 complete image/odometry datasets were collected. When run over these datasets, the visual tracker succeeded in maintaining target lock through 10 complete sequences. Primary failure modes were due to abrupt intensity changes because of indoor lighting or rover shadow. All but one of the failures were corrected by simply re-running the visual tracker with a more appropriate intensity threshold; in the final failed sequence the target was the same color as the back-

[1] http://marscam.jpl.nasa.gov/

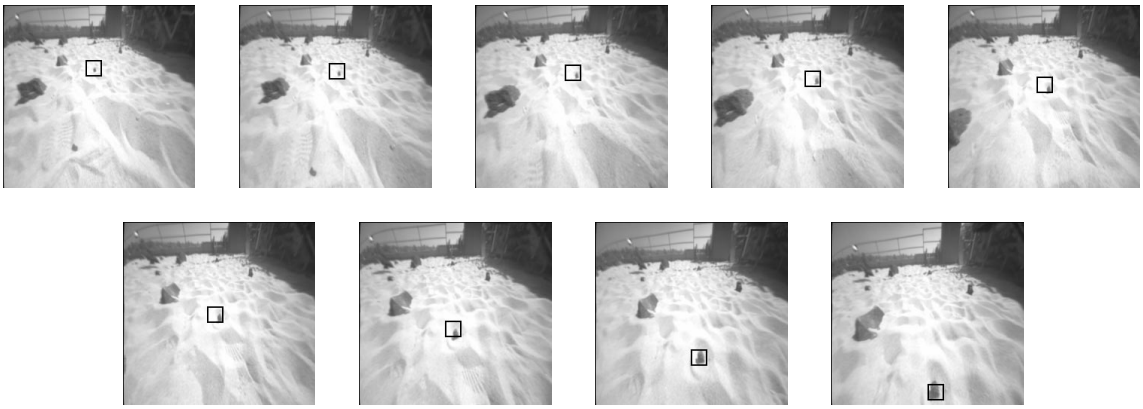

Figure 4: Sample tracking sequence.

ground.

In general, failures can occur when:

- The target leaves the camera FOV, so no range data is available and tracking depends entirely upon noisy odometry.
- The target is visible, but no range data is computed. This can happen if the stereo optics are not properly set for current lighting conditions.
- Multiple targets are visible in the search window and odometry is poor. Additional filtering based on range data could alleviate this, as could matching based on more than a single shape feature (i.e., not just the elevation maximum).
- The target is visible but outside the search window. This happens when the rover climbs over very hilly terrain, if the pose is not measured and used to predict the search window starting point. One could search again using revised motion parameters, or improve the pose sensing.
- Tracking is fine, but the rock is not picked up. This can occur if the rover gets stuck in a servoing loop, attempting to make small changes in position. On sandy soil, such maneuvering introduces much positional uncertainly.
- The target is the same color as the background, so the intensity filter is irrelevant or misleading.

4.1 Mast Placement

This algorithm was also applied successfully to the placement of *Rocky 7's* flexible mast arm on a rock outcropping. The limited degrees of freedom in *Rocky 7's* mast dictate that the vehicle must face the target point's tangent plane on the surface of a boulder to enable complete coverage by the end-effector. For general targets (anywhere on the surface of a boulder) the surface normal is computed from the range data at closest approach, and a two-arc trajectory generated to ensure that the vehicle approaches the rock normal to the tangent plane of the target. However, since this algorithm servos on the local elevation maximum, only targets on the tops of rocks were able to be specified.

During several trials in the Mars Yard *Rocky 7* successfully tracked targets (the tops of boulders 20-50 cm tall) over 5 meters away using the 43-degree FOV stereo cameras in the mast head and successfully placed the end effector on the target. For this application Target Reacquisition occurred after every 50 cm of motion. Execution of the entire sequence (Target Selection, $8-10$ iterations of Target Reacquisition, and successful Mast Placement following the two-arc path generation) typically completed within four minutes when the target was just over 5 meters away.

5 Future Work

In the future we hope to reduce our dependence on the brightness-based filter by matching the entire shape of the terrain around the target (not just its peak) using the technique of [Ols99], and by improving the position and pose estimates using visual feature tracking on the whole scene using a technique from [Mat89]. These improvements should allow tracking of targets anywhere on a rock, enabling a more general mast placement capability, and should also enable tracking of targets that leave the field of view. We would also like to be able to specify multiple targets in a single image, and enable the rover

to keep track of (and acquire) them accurately even if they leave the field of view of the cameras.

6 Acknowledgements

We are grateful to Long Range Science Rover (LRSR) task for providing the *Rocky 7* rover and time to use it, and to the other members of the LRSR team for their work in developing and supporting the vehicle, especially Samad Hayati, Rich Volpe, Bob Balaram, Bob Ivlev, Sharon Laubach, Alex Martin-Alvarez, Larry Matthies, Clark Olson, Rich Petras, Rob Steele, and Yalin Xiong. The work described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract to the National Aeronautics and Space Administration.

References

[Bal99] J. Balaram. Kinematic state estimation for a mars rover. *To appear in special issue of* Robotica *on Intelligent Autonomous Vehicles*, 1999.

[HGT95] Gregory D. Hager, Gerhard Grunwald, and Kentaro Toyama. *Intelligent Robotic Systems*, chapter Feature-Based Visual Servoing and its Application to Telerobotics. Elsevier, Amsterdam, 1995. V. Graefe, editor.

[Mat89] Larry Matthies. *Dynamic Stereo Vision*. PhD thesis, Carnegie Mellon University Computer Science Department, October 1989. CMU-CS-89-195.

[Nis90] H. K. Nishihara. Real-time implementation of a sign-correlation algorithm for image-matching. Technical report, Teleos Research, February 1990.

[Ols99] Clark F. Olson. Subpixel localization and uncertainty estimation using occupancy grids. In *International Conference on Robotics and Automation*, pages 1987–1992, 1999. http://robotics.jpl.nasa.gov/people/olson/abstracts/icra99.html.

[PAW+98] L. Pedersen, D. Apostolopoulos, W. Whittaker, T. Roush, and G. Benedix. Sensing and data classification for robotic meteorite search. In *SPIE Photonics East*, Boston, 1998.

[PK93] N. Papanikolopoulos and P.K. Khosla. Adaptive robotic visual tracking: theory and experiments. *IEEE Transactions on Automatic Control*, 38:1249–1254, March 1993.

[THM+96] D.A. Theobald, W.J. Hong, A. Madhani, B. Hoffman, G. Niemeyer, L. Cadapan, J.J.-E. Slotine, and J.K. Salisbury. Autonomous rock acquisition. In *AIAA Forum on Advanced Development in Space Robotics*, Madison, WI, August 1996.

[Vol99] Richard Volpe. Navigation results from desert field tests of the Rocky 7 mars rover prototype. *International Journal of Robotics Research*, Accepted for Publication, Special Issue on Field and Service Robots 1999. http://robotics.jpl.nasa.gov/people/volpe/papers/JnavMay.pdf.

[WSN85] L. E. Weiss, A. C. Sanderson, and C. P. Neuman. Dynamic visual servo control of robots: an adaptive image-based approach. In *International Conference on Robotics and Automation*, pages 662–668, March 1985.

[WTB97] David Wettergreen, Hans Thomas, and Maria Bualat. Initial results from vision-based control of the Ames Marsokhod rover. In *IEEE International Conference on Intelligent Robots and Systems*, pages 1377–1382, Grenoble, France, September 1997. http://img.arc.nasa.gov/ papers/iros97.pdf.

[XM97] Yalin Xiong and Larry Matthies. Error analysis of a real-time stereo system. In *Computer Vision and Pattern Recognition*, pages 1087–1093, 1997. http://www.cs.cmu.edu/~yx/papers/StereoError97.pdf.