Learning to Behave: Adaptive Behavior for Planetary Surface Rovers

Terry Huntsberger
Jet Propulsion Laboratory
MS 82-105, 4800 Oak Grove Drive
Pasadena, CA 91109
Terry.Huntsberger@jpl.nasa.gov

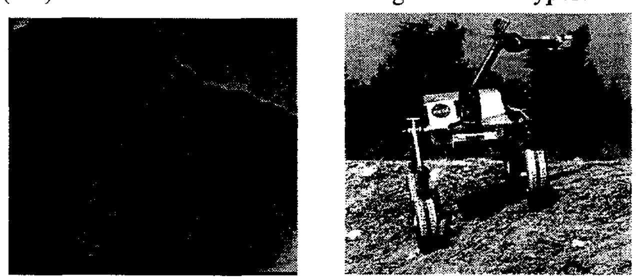
Hrand Aghazarian
Jet Propulsion Laboratory
MS 82-105, 4800 Oak Grove Drive
Pasadena, CA 91109
Hrand.Aghazarian@jpl.nasa.gov

Abstract

Robotic missions to planetary surfaces are becoming more ambitious and of longer duration. The nominal mission timeline for the MER (Mars Exploration Rover) called Spirit currently on the Martian surface is 90 days, with extensions to 180 days depending on rover health. The upcoming 2009 MSL (Mars Science Laboratory) mission is planned to be 300-500 days and will possibly involve traverses on the order of a kilometer or more. Due to time delays of up to 40 minutes round-trip for control, the rovers must have a high degree of onboard autonomous behavior that must also adapt to changing health and environmental conditions during a long duration mission. This paper presents an algorithm for onboard adaptive learning of weights within a rover hierarchical behavior control framework called SMART (System for Mobility and Access to Rough Terrain). SMART is based on earlier work in free flow behavior hierarchies for planetary surface rovers (Huntsberger & Rose, 1998; Huntsberger, 2001). We also present the results of some preliminary laboratory and field studies.

1. Introduction

The Decadal Report (Sykes, 2002) designates autonomy and hard-to-reach access as mobility technology developments for rover exploration of lunar and planetary surfaces, with autonomy cited as high priority. These technology developments are key to surface mission risk mitigation, cost containment for ground operations, as well as high-value science on rough terrain (example shown in Figure 1(a)) beyond the capabilities of current rover designs. The JPL technology prototype rover SRR (Sample Return Rover) shown in Figure 1(b) has the ability to adapt itself to changing terrain. Advances in adaptive behavior techniques tailored to robotic systems that can adapt to rapidly changing terrain through the blend of intelligent sensing (IS) and reflexive behaviors (RB) will enable access to a wide range of terrain types.



(a) (b)

Figure 1: Planetary surface terrain and technology example for autonomous access to high risk, scientifically interesting regions. (a) Mars cliff-face with signs of water outflows; (b) JPL technology prototype of a terrain-adaptive reconfigurable rover.

An optimal solution to this blending process is not achievable in practice due to uncertainty in the sensing and prediction capabilities of the rover. The behavior-based approach to the blending problem used in our previous work is based on multi-objective decision making techniques (Pirjanian, 1998; Pirjanian and Mataric, 2000; Schenker, *et al.*, 2003a-b). Although the state-of-the-art for motion planning was advanced under this approach, integration of other mission autonomy components such as goal achievement and resource management is complex and is difficult to verify. We have developed a unified coherent framework called SMART (System for Mobility and Access to Rough Terrain) to solve the blending process at the system level by treating rover motion, rover health, and resource management as a free flow behavior hierarchy (Huntsberger & Rose, 1998; Huntsberger, 2001). The unique aspect of our approach is enabling the rover to recognize adverse terrain conditions beyond its optimal operational envelope, and intelligently adapting driving techniques (e.g. crabbing motion as a switchback method for steep terrain ascent) and/or rover geometry for safe traverses (Schenker, *et al.*, 2003a-b).

We previously developed a control architecture called BISMARC (Biologically Inspired System for Map-based Autonomous Rover Control) for long duration missions (Huntsberger and Rose, 1998; Huntsberger, 2001). It is based on a modified free-flow hierarchy (FFH) similar to the DAMN architecture (Rosenblatt and Payton, 1989; Tyrrell, 1993), and has been used successfully for a number of different simulated mission scenarios including multiple cache retrieval (Huntsberger, 1997), fault tolerance for long duration missions (Huntsberger, 1998), and site preparation (Huntsberger, *et al.*, 1999). The system (high level overview shown in Figure 2) includes all aspects of safety, self-maintenance, and goal achievement that robotic systems require for a sustained planetary surface presence. More details concerning BISMARC are found in the next section. Its original implementation used fixed weights in the FFH, and as such, was not able to adapt to environments outside of the original world model. BISMARC, with the added learning mechanisms, serves as the core action generation and action selection portions of SMART.

The research issues that must be addressed to enable safe, adaptive rover mobility in highly sloped areas include (1) determination of near-optimal strategies (in terms of rover safety) for adaptive rover pose reconfiguration and driving that are computationally feasible for onboard implementation, (2) determination of a representation for uncertainty in sensing and prediction of rover and environment state, and (3) determination of resource management strategies that mitigate risks such as loss of battery power and/or drive motors. These issues must be addressed within the onboard rover computational capabilities that are limited due to power constraints. The MER CPU is a RAD6000 processor running at 27Mhz and the CPU planned for MSL will be based on a radiation-hardened version of the PowerPC750 running at 200Hz, as compared to current Pentium speeds of 2Ghz. The behavior of the system must also be adaptive since the environmental conditions on a planetary surface such as Mars are harsh and the terrain has a great deal of variability.

Prior research by Tyrrell (Tyrrell, 1993) and Bryson (Bryson, 2000) demonstrated superior performance of a hierarchical system for action selection over purely reactive systems. In particular, the agents in the Edmund system of Bryson (Bryson, 2000) are built as related sensing and action functions that exhibit selective attention with the payoff of a higher efficiency than the modified Rosenblatt and Payton (RP) mechanisms of Tyrell (Tyrrell, 1993). A comprehensive overview of action selection systems can be found in Bryson (Bryson, 2001). Although BISMARC uses the modified RP mechanisms, the nodes in the free flow behavior hierarchy perform operations that are more sophisticated than simple

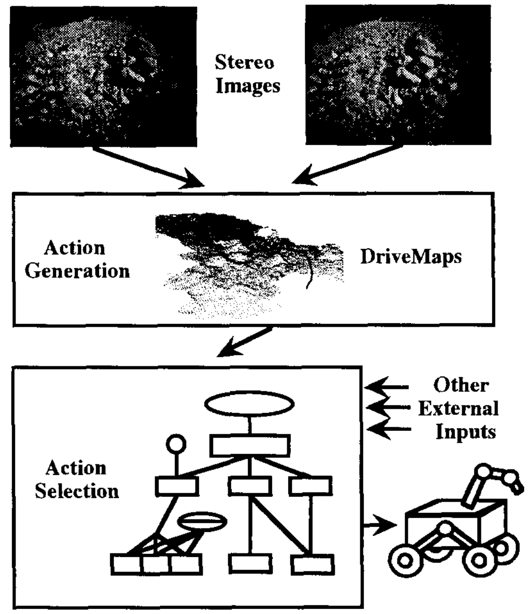


Figure 2: Two level BISMARC architecture with stereo processing, action generation, and action selection subsystems.

combination. In some sense, they are closer to the *competence* structures of Bryson (Bryson, 2000), in that a collection of plan elements are organized as a prioritized finite state machine whose outputs converge on a specific goal. These nodes have undergone extensive evaluation at the modular level either through field or mission testing.

To date, there has been very little research into learning for hierarchical action selection systems which are typically characterized by multiple, possibly conflicting goals. The dominant learning strategy for single goal achievement such as robotic navigation has been reinforcement learning (RL), an unsupervised method that seeks to maximize a reward signal based on the utility of pairings of input and output states and their subsequent actions (Kaelbling, 1993; Sutton and Barto, 1998). One of the most popular RL algorithms is Q-learning (Watkins, 1989) and its variations such as Q-PSP (Horiuchi, *et al.*, 1996), and hierarchical Q-learning (Lin, 1993). RL algorithms typically suffer from slow convergence, large state spaces, and difficulties in handling uncertain sensory inputs. These concerns were addressed for deterministic environments using a forgetting mechanism in a penalty-based hierarchical Q-learning algorithm, which reduces the amount of state information that an agent must maintain by using a low level agent to maintain local state information and a high level agent to maintain global state information (Yen, *et al.*, 2001; Yen and Hickey, 2002). During planetary surface rover operations, the prediction of a state following an action is difficult since it is closer to a non-deterministic process due to interactions with the terrain.

Most of the RL studies to date have been confined to simulations and interior navigation in 2-D environments.

An alternate learning system that performs in the presence of a multiple conflicting goals where subtasks are only partially satisfied (Maes, 1991) is W-learning and its variations, which are based on compromise or negotiated decision making between agents (Humphrys, 1997). W-learning is a memory efficient method that is more suited for operation onboard planetary surface rovers than traditional or hierarchical Q-learning systems, and a temporally prioritized modification of it is running under SMART.

The next section briefly describes the organization of BISMARC, followed by a discussion of the learning mechanism of the system. We close with experimental studies and conclusions.

2. BISMARC Organization

BISMARC is organized as a two level system (shown in Figure 2). The first level generates possible motor actions using stereo images and the second level uses these action hypotheses coupled with external and internal inputs to drive the actuators on the robot. The *DriveMaps* algorithm used for action generation analyzes local range information for clear paths relative to a goal and is currently implemented on a number of technology prototype rovers at JPL (Huntsberger, *et al.*, 2002). A fuzzy adaptive behavior system with similar capabilities to *DriveMaps* is described in (Tunstel, 2001).

Figure 3 shows the action selection hierarchy for a rough terrain navigation mission that is used for the experimental studies reported in this paper. The rectangular boxes represent behaviors and the ovals are sensory inputs (either fixed, direct, or derived). At the top are the high level behaviors including *Don't Tip Over, Go to Goal, Avoid Obstacles, Preserve Motors, Warm Up, Get Power,* and *Sleep at Night*. These goals are related to both task and rover safety. For example, since most planetary surface rovers have only visual sensors for navigation, the sensory input for *Proximity to Night* is derived from knowledge of the sun's position and forces the rover to sleep at night by weighting the input to *Sleep at Night* heavier (4.0) than any other behavior in the hierarchy. The *Avoid Obstacles* behavior uses the output of the *DriveMaps* algorithm as recommendations for viable paths. The rovers are equipped with solar panels and the *Rest* behavior allows the batteries to recharge if the sun is up. The *Rest* behavior is also used to cool down the motors for *Preserve Motors* if they are working too hard going up a steep slope, or to stop and turn on the heaters for *Warm Up* if the internal temperature of the rover drops below a safety threshold.

The intermediate level *Change CG* behavior is an example of a sophisticated combination behavior discussed in Section 1 that works to shift the center of gravity of the rover (see Figure 1(b)) much like an animal does in response to traveling up or along a steep slope. This behavior is implemented using a finite state machine based on a well-tested algorithm for pose reconfiguration (Schenker, *et al.*, 2003a-b). The algorithm uses the onboard gyroscopes and accelerometers, which would be equivalent to the inner ear mechanism in mammals for roll and pitch determination. Recommendations for shoulder

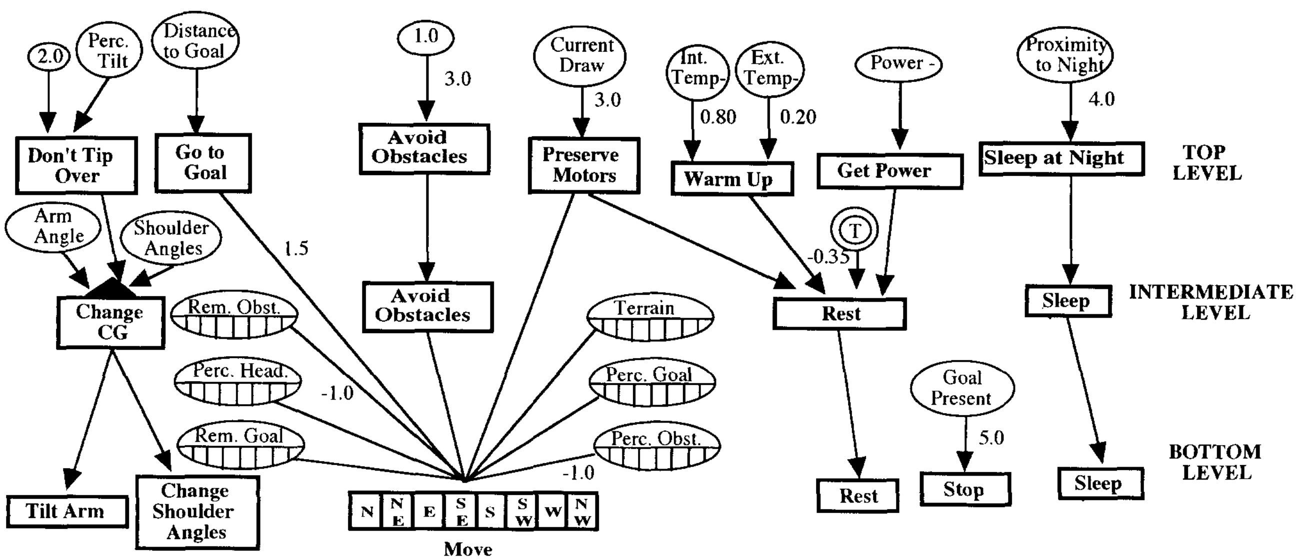


Figure 3: Free-flow hierarchy action selection mechanism for rough terrain navigation mission scenario. Ovals represent inputs derived from sensory stimuli, rectangular boxes are behaviors, and double ovals are temporal penalties. All weights on inputs to behaviors are 1.0 unless otherwise noted. Segmented boxes and ovals represent directional inputs (only cardinal directions shown but in practice continuous coverage). See text for further details.

angle and arm end position changes to help stabilize the rover are generated and passed on to the bottom level behaviors.

The intermediate level behaviors are designed to interact with both the short term memory (STM), which corresponds to perceived sensory stimuli, and the long term memory (LTM), which encodes remembered sensory information. Control loops are prevented through temporal penalties (shown as T-ovals in Figure 3) that constrain the system to only repeat a behavior a predetermined number of times. The bottom level behaviors in the hierarchy fuse the sensory inputs and the activations of the higher level behaviors in order to select appropriate actions for rover safety and goal achievement. The rover will continue to move until it achieves the goal position as determined by a rover localization algorithm (Hoffman, *et al.*, 1998) shown as the *Goal Present* input to *Stop* in Figure 3, or its health deteriorates due to dead batteries, freezing, burned out motors, or tipping over.

The STM (Short Term Memory) used in SMART is distributed in the sense that a detailed occupancy grid (5 cm resolution) (Elfes, 1987) is kept by *DriveMaps* of the last 12.5 m of travel, in addition to the previous and current rover-state vectors (shown as ovals in Figure 3). These sensor readings are weighted by a perceptual uncertainty based on the absolute difference between the related portions of the previous and current state vectors.

SMART's map-based LTM (Long Term Memory) is similar to that of hippocampus place cells. Landmarks corresponding to obstacles and goals are extensively mapped and stored for comparison to perceived inputs, with a probabilistic update of memories based on the positional variance of the rover and the match strength of the current perception to memory contents. A LTM landmark is encoded as a four-byte field that includes relative height of the landmark (2 bytes), actions leading to the landmark (1 byte), and accelerometer readings on the robot (1 byte). This approach is similar to the coupled goal/representation approach of (Mataric, 1992; Mataric, 1997) and saves on-board memory use. An alternative approach is an occupancy grid that gives dense coverage of the environment, but doesn't scale well for long duration planetary surface missions.

3. Learning Mechanism

Learning mechanisms for rovers on a planetary surface require low computational overhead, reactivity even in uncertain environments, no loss of internal state information, combination of possibly conflicting behaviors, and the localization of sensory input to the appropriate modules. The learning algorithm used in SMART includes all of these properties.

Combining the inputs to a behavioral node is usually calculated as a simple weighted summation. This approach leads to potential problems in the case where the same goal triggers two or more behaviors and the utility of a behavior lower in the hierarchy should not be the sum of their activations. For example, in Figure 3, the *Preserve Motors* goal feeds into the *Rest* and the *Move* behaviors. The *Rest* behavior has a much greater chance of satisfying the goal by keeping the rover alive, since it doesn't involve further load on already overloaded motors. A better activation function, used in BISMARC, balances strong preferences as well as aversion behaviors (i.e., *Avoid Obstacles*) (Tyrrell, 1993). The activation output of the behaviors associated with directional sensory inputs (shown as segmented ovals in Figure 3) are multiplied by the sensor stimuli before being used by lower levels in order to suppress activation in directions other than that favored by the sensors.

The weights on the links between modules are usually heuristically determined based on mission goals. These goals are specified at a relatively high level without complete knowledge of the operating environment of the rover. In addition, rover health will degrade as the mission progresses, and weights chosen at full health may no longer be appropriate. A modified version of the W-learning algorithm of Humphrys (Humphrys, 1997) is the approach used in SMART to address this problem. Weights are adjusted using the maximize collective happiness (MCH) algorithm (Humphrys, 1997) applied to sub-paths in the hierarchy.

The *Tilt Arm, Change Shoulder Angles,* and *Move* actions at the lowest level in the FFH shown in Figure 3 can be done simultaneously. However, progress towards the goal will be compromised if the rover tips over, so there is a dynamic relationship between the two higher level goals of *Go to Goal* and *Don't Tip Over*. The MCH algorithm is applied to both of the sub-paths in the hierarchy with a time delay between update of the weights. The *Don't Tip Over* sub-path weights are changed in the first time slice, followed by the *Go to Goal* sub-path weights. This maintains the rover health, while at the same time making progress towards the goal. Another instance where this process is applied is the relative direction that the rover moves. In order to *Preserve the Motors*, the rover will attempt to climb a steep incline, and either back off or rest if the perceived motor currents in the rear wheels are too high. If the weights are not dynamically adjusted, this could lead to dithering where the rover attempts to climb, backs off, and then attempts to climb in the same direction. Adaptive weighting using the MCH algorithm changes the direction of attack, since progress towards the goal is being compromised by the dithering. For this situation, there is a time delay between application of MCH to the two sub-paths of *Go to Goal* and *Preserve Motors*, with *Preserve Motors* occurring first, followed by *Go to Goal*.

4. Experimental Studies

In order to determine the utility of SMART for planetary surface operations in rough terrain, we have run three different types of experimental studies: (1) 2000 simulated rough terrain navigation missions, (2) 50 laboratory sequences with SRR, and (3) 4 sequences with SRR in natural terrain in the Arroyo Seco outside JPL. We have attempted to match the fidelity of the simulation models for terrain and rovers to those used for the laboratory and field studies.

4.1 Simulation studies

The first series of experimental studies used simulated terrain based on MOLA (Mars Orbiter Laser Altimeter) data from the Dao Valis region of Mars, which had slopes of up to 65°. A 200 meter by 200 meter sub-area of the rough terrain dataset is shown in Figure 4 and a view of the SRR during one of the simulation runs is shown in Figure 5. Mission success was defined as the attainment of the randomly selected goal position without dying due to freezing, dead batteries, burned out motors, or tipping over. The experimental setup included:

- Random starting and goal positions
- Timestep of 0.1s
- 10% loss of traction in rocky terrain
- 1 sq. km study area (5 cm resolution)
- Top speed of 15 cm/sec

The model of SRR matches the physical platform and has two sets of stereo cameras, one body-mounted and one mast mounted, a 3 DOF (degrees of freedom) manipulator and a twelve week battery lifetime supplemented with solar panels.

Our studies had a 95.9% mission success with the onboard adaptive learning mechanism, and a 43% success rate without the adaptive learning. The primary failure mode (3.8%) for the system with learning enabled was

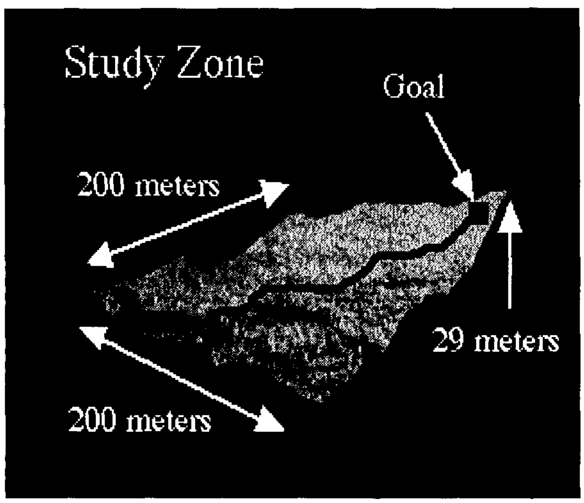


Figure 4: Terrain model for rough terrain navigation mission, with goal position at box and path shown as solid line. Study zone is 200 m by 200 m and terrain variation is from -271 m to 300 m.

Figure 5: The SRR climbing a 35° slope in simulated terrain derived from MOLA data in the Dao Valis region of Mars. The model of the rover contained full kinematics and dynamics and used a probabilistic slip assumption. The FFH shown in Figure 2 was used for control and adaptive learning for 2000 simulation runs.

dead batteries which from a mission standpoint would indicate a need for larger solar panels. An analysis of the 57% of the missions that failed with no learning enabled gives:

- Tipping over - 27%
- Dead batteries - 15%
- Burned out motors - 9%
- Freezing - 6%

Since 27% of the missions failed due to tipping over, the initial weights for inputs to *Move* were set too high, giving an overall bias to the *Get to Goal* behavior over rover safety related behaviors such as *Don't Tip Over*.

4.2 Laboratory studies

The second set of experimental studies was run in the Planetary Robotics Lab (PRL) at JPL and used the JPL technology prototype rover SRR shown in Figure 6. SRR has independently articulated shoulders which allow it to dynamically change its pose and lean much like an animal does on sloped terrain. The full range of shoulder movement is shown in Figure 6. SRR also has independent four wheel drive and independent four wheel steering enabling it to travel sideways.

One of the experimental runs is shown in Figure 7, where we have set up a worse case scenario of opposing hills and valleys for the rover. The SRR (Sample Return Rover) successfully negotiated the course based on a subnet of the full hierarchy shown in Figure 3. This subnet included the *Don't Tip Over, Go to Goal, Avoid Obstacles*, and *Preserve Motors* top level nodes. The *Warm Up, Get Power,* and *Sleep at Night* top level node activation levels

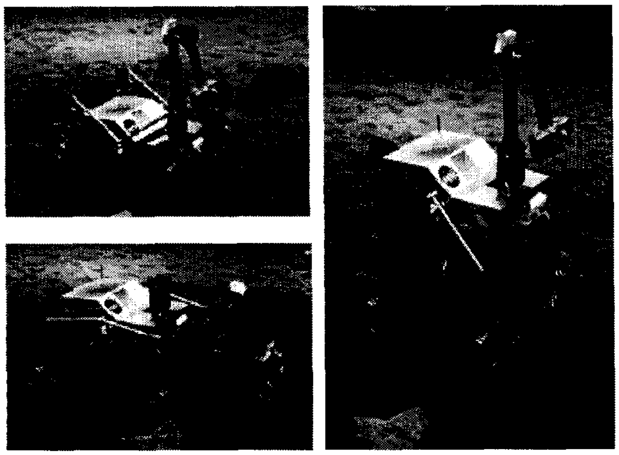


Figure 6: Sample Return Rover (SRR) range of hardware adaptation including clockwise from upper left - the lowest range of the shoulder articulation, the highest range of shoulder articulation, and the mid-range of shoulder articulation coupled with extended arm movement.

were all set to zero since the interior of the lab was warm and not exposed to the sun.

Another series of laboratory runs used a ramp set at a 65° slope with the rover positioned at the bottom. The goal position was given as the top of the ramp which is beyond SRR's stability level even with shoulder reconfiguration. Initially the rover attempted to climb the slope, but repeatedly backed off and then tried again. This behavior can be traced to the combination of *Go to Goal* and *Preserve Motors* using the default weights. The learning algorithm progressively reduced the *Go to Goal* weight from 1.5 to 0.5 which caused the rover to try to skirt the ramp by moving sideways while still maintaining movement towards the goal. As the rover cleared the side of the ramp it then started movement towards the goal due to the learning algorithm increasing the weight on the *Go to Goal* behavior output. The learning algorithm is providing SMART with the adaptability that is needed for an unknown environment. Further testing will be needed to verify that rover safety is not compromised. In the field, this behavior would be equivalent to the rover trying to find a safe way up a slope to get to the goal.

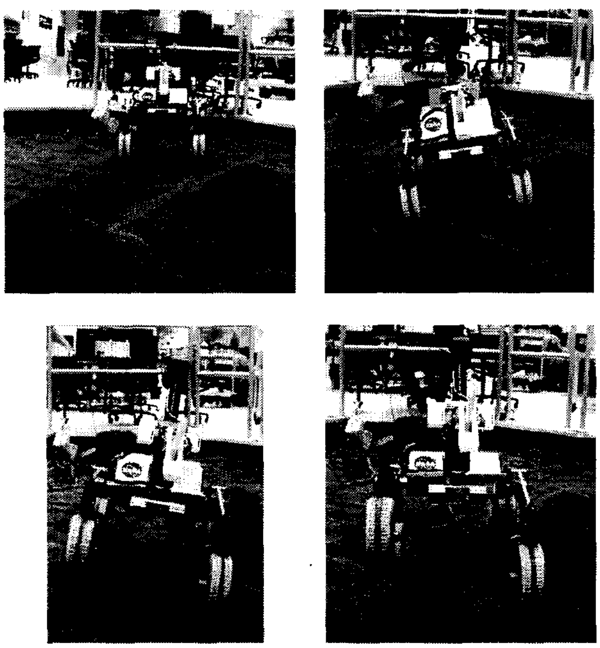


Figure 7: Clockwise from upper left: SRR performing continuous pose reconfiguration using its adjustable shoulders during a traverse in the Planetary Robotics Lab at JPL. The terrain was a set of two opposing hills and valleys, with 45° degree slopes.

4.3 Field studies

The last series of experimental studies was done in the Arroyo Seco, a dry wash that is next to JPL. This site is used for technology prototype rover testing and is characterized by a mixture of benign sand and rocky beds that have been scoured by the periodic water flow bordered by steeply sloped cliffs. An example of the terrain with SRR during a traverse is shown in Figure 8. The learning component of SMART was not fully implemented at the time, so only qualitative results are available at this time.

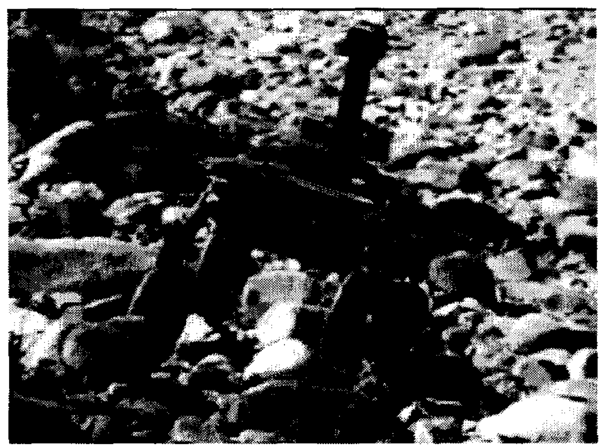


Figure 8: SRR in the bottom of a rock-strewn gully in the Arroyo Seco outside of JPL. The right shoulder is almost horizontal compared to the left one because the rover just came off of the rock behind the right rear wheel.

We were only able to complete a preliminary series of 4 runs in the Arroyo Seco and will return for more data collection in the spring of 2004 after the winter rains. An example of the skirting behavior along a slope, as previously seen in the laboratory studies discussed in section 4.2, is shown in Figure 9, where the rover approaches the slope in the left frame and is not able to climb, skirts to the side in the middle frame, and finally

Figure 9: Skirting behavior of SRR along the length of a slope in the Arroyo Seco wash outside of JPL where the rover initially can not get enough traction to climb so the direction of travel favors lateral motion. Time flows from left to right in the sequence with the left frame being the initial approach almost parallel to the slope, the middle frame showing a change in rover heading more perpendicular to the slope for better traction on the top of the rise, and the right frame showing the rover continuing along its initial heading towards the goal. The mast was fixed in its orientation for this run and would have given the rover more traction if pointed up slope.

gets enough traction to climb to the top of the rise and continue on towards the goal.

5. Conclusions

We have developed an autonomous rover control system called SMART for planetary rovers traversing rough and highly sloped terrain. It is based on the previously developed free flow hierarchical action selection of BISMARC, coupled with an onboard learning mechanism for changing weights in the hierarchy. The learning mechanism enabled SMART to maintain rover health in both simulated and actual rover studies in rough terrain. Of particular importance for future NASA rover missions was the analysis of the rover failures, indicating that an additional 52.9% of missions would potentially be successful with adaptive learning. Our current directions include integration of the SMART control techniques into the recently developed CAMPOUT (Control Architecture for Multi-robot Planetary Outposts) running on two technology prototype rovers at JPL (Huntsberger, *et al.*, 2003; Schenker, *et al.*, 2003a).

6. Acknowledgments

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. The views contained herein are those of the authors and are not meant to in any way imply that they are those of NASA or JPL.

References

Bryson, J.J. (2000). Hierarchy and Sequence vs. Full Parallelism in Action Selection. In *From Animals to Animats 6, Proc. of the Sixth Intern. Conf. on Simulation of Adaptive Behavior (SAB '00)*, 147-156.

Bryson, J.J. (2001). *Intelligence by Design*. Ph.D. dissertation, Dept. of Elec. Engineering and Computer Science, MIT, Cambridge, MA, USA.

Elfes, A. (1987). Sonar-based real-world mapping and navigation. *IEEE Journal of Robotics and Automation*, RA-3(3): 249-265.

Hoffman, B.D., Baumgartner, E.T., Huntsberger, T., and Schenker, P.S. (1999). Improved rover state estimation in challenging terrain. *Autonomous Robots*, 6(2): 113-130.

Horiuchi, T., Fujino, A., Katai, O., and Sawaragi, T. (1996). Q-PSP Learning: An exploitation-oriented Q-learning algorithm and its applications. In *Proc. IEEE International Sympos. on Evolutionary Computation*, 76-81.

Humphrys, M. (1997). *Action Selection Methods using Reinforcement Learning*. PhD thesis, University of Cambridge, Cambridge, UK.

Huntsberger, T.L. (1997). Autonomous multi-rover system for complex planetary surface retrieval operations. In *Proc. Sensor Fusion and Decentralized Control in Autonomous Robotic Systems*, SPIE Vol. 3209, 220-229.

Huntsberger, T.L. (1998). Fault-tolerant action selection for planetary rover control. In *Proc. Sensor Fusion and Decentralized Control in Robotic Systems*, SPIE Vol. 3523, 150-156.

Huntsberger, T.L. (2001). Biologically inspired autonomous rover control. *Autonomous Robots*, 11(11): 341-346.

Huntsberger, T.L., Mataric, M.J., and Pirjanian, P. (1999). Action selection within the context of a robotic colony. In *Proc. Sensor Fusion and Decentralized Control in Robotic Systems II*, SPIE Vol. 3839, 84-91.

Huntsberger, T., Aghazarian, H., Cheng, Y., Baumgartner, E.T., Tunstel, E., Leger, C., Trebi-Ollennu, A., and Schenker, P.S. (2002). Rover Autonomy for Long Range Navigation and Science Data Acquisition on Planetary Surfaces. In *Proc. 2002 IEEE International*

Conf. on Robotics and Automation (ICRA2002), Washington, DC, 3161-3168.

Huntsberger, T., Pirjanian, P., Trebi-Ollennu, A., Nayar, H.D., Aghazarian, H., Ganino, A., Garrett, M., Joshi, S.S., and Schenker, P.S. (2003). CAMPOUT: A Control Architecture for Tightly Coupled Coordination of Multi-Robot Systems for Planetary Surface Exploration. *IEEE Trans. Systems, Man & Cybernetics, Part A: Systems and Humans, Special Issue on Collective Intelligence*, 33(5): 550-559.

Huntsberger, T.L. and Rose, J. (1998). BISMARC. *Neural Networks*, 11(7/8): 1497-1510.

Kaelbling, L.P. (1993). *Learning in Embedded Systems*. MIT Press, Cambridge, MA, USA.

Lin, L.-J. (1993) *Reinforcement Learning for Robots Using Neural Networks*. Ph.D. dissertation, Carnegie Mellon University, Pittsburgh, PA, USA.

Maes, P. (Ed.) (1991). *Designing Autonomous Agents Theory and Practice from Biology to Engineering and Back*, MIT Press.

Mataric, M.J. (1992). Integration of representation into goal-driven behavior-based robots. *IEEE Trans. on Robotics and Automation*, 8(3): 304-312.

Mataric, M.J. (1997). Behavior-based control: Examples from navigation, learning, and group behavior. *Journal of Experimental and Theoretical Artificial Intelligence, Special Issue on Software Architectures for Physical Agents*, 9(2-3): 323-336.

Pirjanian, P. (1998). *Multiple Objective Action Selection and Behavior Fusion Using Voting*. PhD dissertation, Laboratory of Image Analysis, Department of Medical Informatics and Image Analysis, Aalborg University, Denmark.

Pirjanian, P. and Mataric, M., (2000). Multiple objective vs. fuzzy behavior coordination. In *Lecture Notes in Computer Science on Fuzzy Logic Techniques for Autonomous Vehicle Navigation*, D. Drainkov and A. Saffiotti, (Eds.), Springer-Verlag, 235-253.

Rosenblatt, J.K. and Payton, D.W. (1989). A fine-grained alternative to the subsumption robot control. In *Proc. IEEE/INNS Joint Conf. on Neural Networks*, 317-324.

Schenker, P.S., Huntsberger, T.L., Pirjanian, P., Baumgartner, E.T., and Tunstel, E. (2003a). Planetary Rover Developments Supporting Mars Exploration, Sample Return and Future Human-Robotic Colonization. *Autonomous Robots*, 14(2/3): 103-126.

Schenker, P.S., Huntsberger, T., Pirjanian, P., Dubowsky, S., Iagnemma, K., and Sujan, V. (2003b). Rovers for Intelligent, Agile Traverse of Challenging Terrain. In *Proc. International Conference on Advanced Robotics (ICAR'03)*, University of Coimba, Portugal, 1683-1692.

Sutton, R.S. and Barto, A.G. (1998). *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA.

Sykes, M.V. (Ed) (2002) *The Future of Solar System Exploration, 2003-2013*. NRC Planetary Decadal Report (http://www.aas.org/~dps/decadal).

Tunstel, E. (2001). Ethology as an inspiration for adaptive behavior synthesis in autonomous planetary rovers. *Autonomous Robots*, 11(11): 333-340.

Tyrrell, T. (1993). The use of hierarchies for action selection. *Journal of Adaptive Behavior*, 1(4).

Watkins, C.J. (1989). *Learning from Delayed Rewards*. Ph.D. dissertation, Cambridge University, Cambridge, UK.

Yen, G., Yang, F., Hickey, T., and Goldstein, M. (2001). Coordination of exploration and exploitation in a dynamic environment. In P*roc. International Joint Conference on Neural Networks (IJCNN '01), Volume 2*, 1014-1018.

Yen, G. and Hickey, T. (2002). Reinforcement learning algorithms for robotic navigation in dynamic environments. In *Proc. of the 2002 Intern. Joint Conf. on Neural Networks (IJCNN '02), Vol. 2*, 1444-1449.