Virtual Research Presentation Conference

UNIFIED PROCESSING FOR ROBOTIC ICY TERRAIN EXPLORATION (UPRITE)

Principal Investigator: Brett Kennedy 347B
Co-Is: Joseph Bowkett 347G (narrator), Aryan Naim 393K, Peyman Tavallali 398J
Program: Sure Terrain Assessment Based on Learning the Environment (STABLE)

Assigned Presentation #RPC-250

Jet Propulsion Laboratory
California Institute of Technology

Problem Description

Current planetary rovers typically use “Ground-in-the-loop” (human operators) to make all but the simplest decisions.

Environments like Mars are relatively well understood, meaning systems can be designed for the expected conditions.

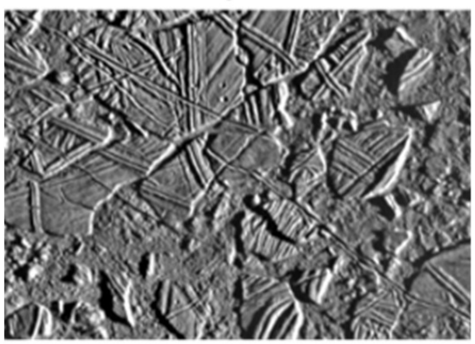
Problem Description

Current planetary rovers typically use “Ground-in-the-loop” (human operators) to make all but the simplest decisions.

Environments like Mars are relatively well understood, meaning systems can be designed for the expected conditions.

Problem: How can we make robotic platforms automatically adapt to environments we know much less about, without asking a human?

Europan Surface



Matanuska Glacier, AK

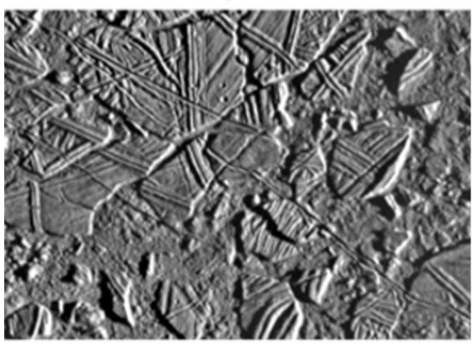
Problem Description

Current planetary rovers typically use “Ground-in-the-loop” (human operators) to make all but the simplest decisions.

Environments like Mars are relatively well understood, meaning systems can be designed for the expected conditions.

Problem: How can we make robotic platforms automatically adapt to environments we know much less about, without asking a human?

Europan Surface



Matanuska Glacier, AK

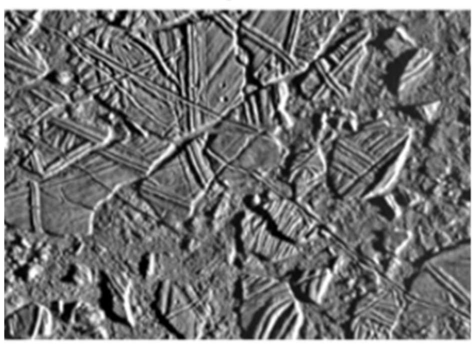
Problem Description

Current planetary rovers typically use “Ground-in-the-loop” (human operators) to make all but the simplest decisions.

Environments like Mars are relatively well understood, meaning systems can be designed for the expected conditions.

Problem: How can we make robotic platforms automatically adapt to environments we know much less about, without asking a human?

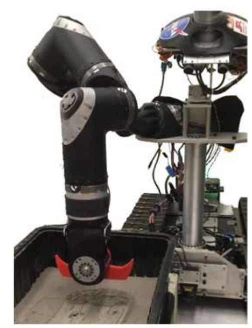
Europan Surface



Matanuska Glacier, AK

Surface Mobility

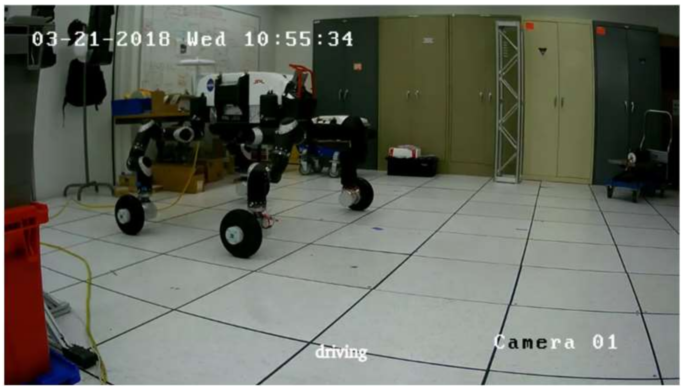
RoboSimian Platform



Excavation

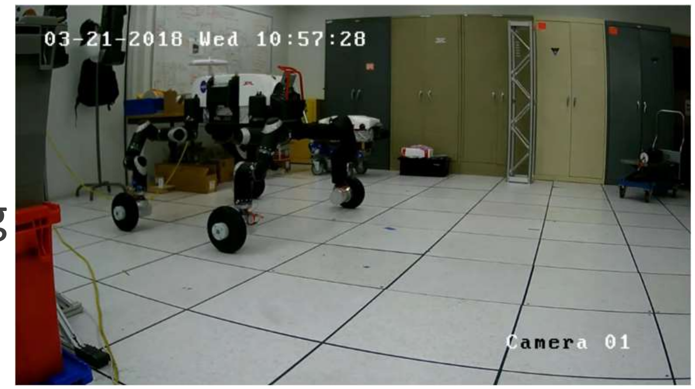
SURROGATE Platform

Use case: Robotic Mobility

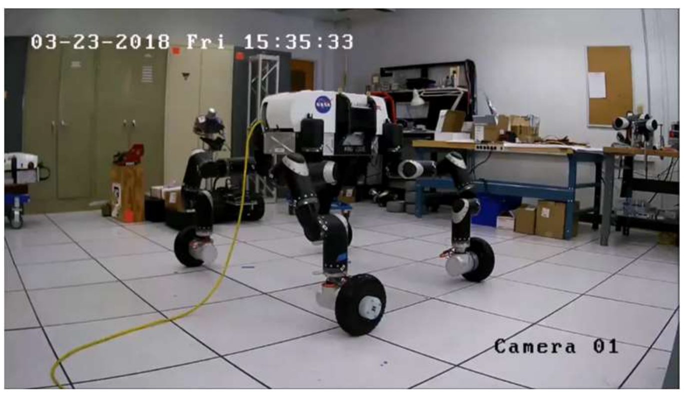


Driving
$\bar{c} = 15\ kJ/m$

Walking

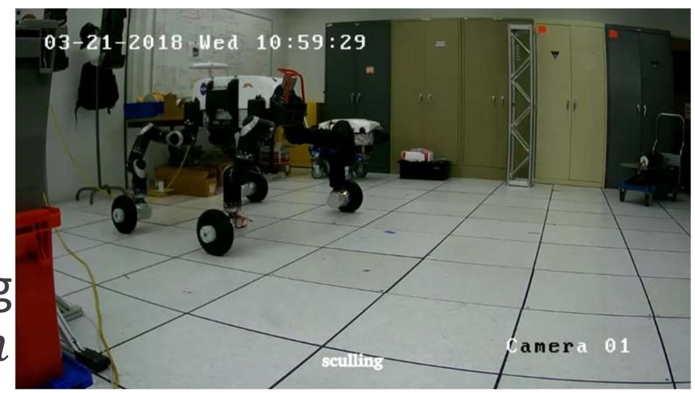


Cost metric: kJ/m



Inchworming
$\bar{c} = 25\ kJ/m$

Sculling
$\bar{c} = 50\ kJ/m$



Use case: Robotic Mobility

- Driving is very efficient, but gets stuck on steep gradients making the energy efficiency go to zero
- Inchworming uses more energy on flat ground, but doesn't get stuck on a sandy slope

Methodology

Approach: Derived from Reinforcement Learning sub-field named Multi-Armed Bandit theory

$$a(t) \in \mathcal{A} := \{1, \ldots, n\} \quad \Longrightarrow \quad r(t)$$

Multi-Armed Bandit = Set of slot machines, or multiple "one-armed bandits"

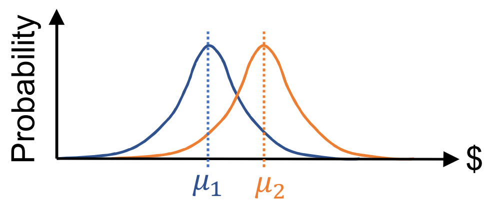
Methodology

Approach: Derived from Reinforcement Learning sub-field named Multi-Armed Bandit theory

$$a(t) \in \mathcal{A} := \{1, \ldots, n\} \quad \Rightarrow \quad r(t)$$

Each option returns a random reward, but each has a different mean or average reward.

$$r(t) \sim X_{a(t)} \qquad \mu_i = \mathbb{E}[\chi_i] \ \forall i \in \mathcal{A}$$



Multi-Armed Bandit = Set of slot machines, or multiple "one-armed bandits"

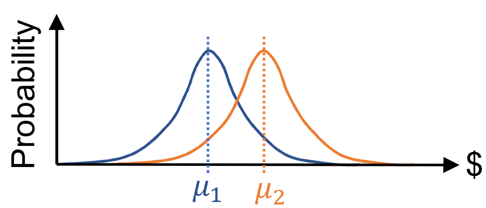
Methodology

Approach: Derived from Reinforcement Learning sub-field named Multi-Armed Bandit theory

$$a(t) \in \mathcal{A} := \{1, \ldots, n\} \quad \longrightarrow \quad r(t)$$

Each option returns a random reward, but each has a different mean or average reward.

$$r(t) \sim X_{a(t)} \qquad \mu_i = \mathbb{E}[\chi_i] \; \forall i \in \mathcal{A}$$



Multi-Armed Bandit = Set of slot machines, or multiple "one-armed bandits"

Problem: Given some number of plays, how do we figure out which machine is best and maximize our winnings?

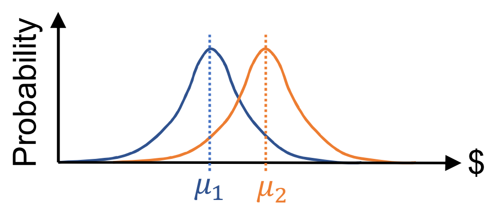
Methodology

Approach: Derived from Reinforcement Learning sub-field named Multi-Armed Bandit theory

$$a(t) \in \mathcal{A} := \{1, \ldots, n\} \quad \Rightarrow \quad r(t)$$

Each option returns a random reward, but each has a different mean or average reward.

$$r(t) \sim X_{a(t)} \qquad \mu_i = \mathbb{E}[\chi_i] \ \forall i \in \mathcal{A}$$



Problem: Given some number of plays, how do we figure out which machine is best and maximize our winnings?

Multi-Armed Bandit = Set of slot machines, or multiple "one-armed bandits"

UPRITE algorithms expand on this for exploration of dynamic environments.

Written into library named RSTAR (Reinforced STochastic AutoRegression)

Methodology

Traditional MAB policies typically test, or *explore*, each of the options for some period then use, or *exploit*, the best one they find

e.g. ϵ_{greedy} , Upper Confidence Bound (UCB)

Distinctions:
- We can determine *a priori* which options we expect to work best

- Our environment might change after we test the options

Methodology

Traditional MAB policies typically test, or *explore*, each of the options for some period then use, or *exploit*, the best one they find

e.g. ϵ_{greedy} , Upper Confidence Bound (UCB)

Distinctions:

- We can determine *a priori* which options we expect to work best

- Our environment might change after we test the options

Naïve ϵ_{greedy} policy learns driving works well on flat ground, but then keeps trying once it gets on a slope and ends up stuck

Use case: Robotic Mobility

UPRITE "decision engine" automatically switches between modes when they become less effective

UPRITE formulation:

Preferential Iterative Update (PIU) policy

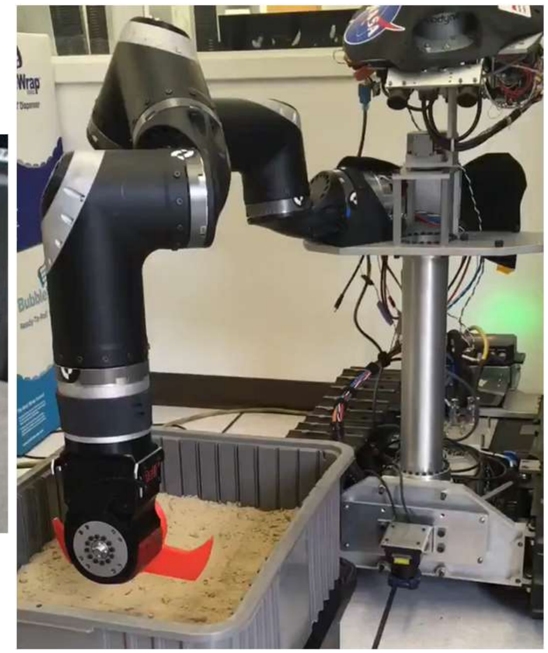
Uses exponential tracking function instead of average:

$$Q_i^{\text{PIU}}(t+1) = Q_i^{\text{PIU}}(t) + \alpha \left[q_i(t) - Q_i^{\text{PIU}}(t) \right] \ \forall i \in \mathcal{A}$$

Slowly "forgets" past measurements

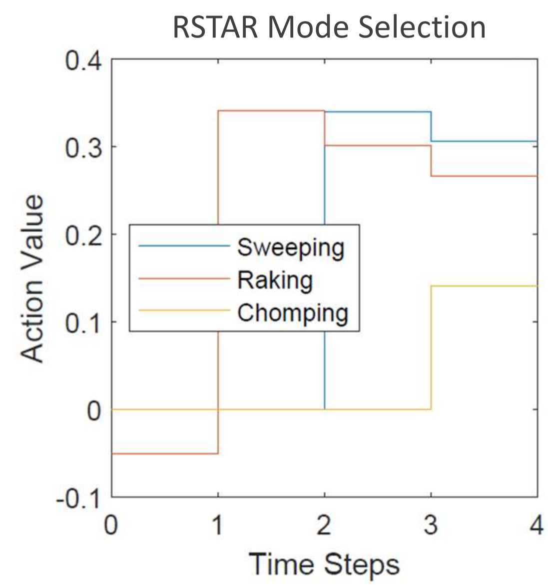
Use Case: Excavation

Sweeping | Raking | Chomping



Cost metric: kJ/cm^3

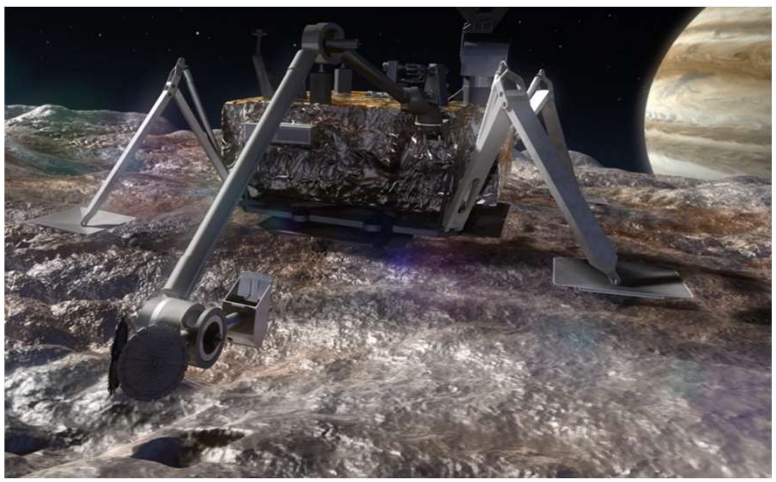
Use Case: Excavation



Algorithm realizes normally better sweeping and raking modes aren't working, switches to more energy intensive chomping

Applications

Any poorly understood environment where human intelligence isn't available to determine the best way to execute a task



Icy Moon Surface Sampling

Matanuska Glacier, AK

Difficult to reach Earth Environments

Publications and References

P. Tavallali, S. Karumanchi, W. Reid, J. Bowkett, B. Kennedy, “A Reinforcement Learning Framework for Space Missions in Unknown Environments”. Presented at *IEEE Aerospace Conference*, 2020.

Aryan Naim, Joseph Bowkett, Sisir Karumanchi, Peyman Tavallali, Brett Kennedy, “Deterministic Iteratively Built KD-Tree with KNN Search for Exact Applications”. Submitted to *IEEE Computer Society* Sept 2020.

Joseph Bowkett, William Reid, Joel Burdick, Peyman Tavallali, Jay Jasper, Blair Emanuel, Brett Kennedy, and Sisir Karumanchi, “The Obedient Multi-Armed Bandit: Selecting Operating Modes in Uncertain Environments with Bounded Exploration”. To be submitted to *International Journal of Robotics Research*.