Zero-Shot Context Identification through Clustering and Foundation Modeling for Friction Estimation

Renukanandan Tumu*, Ahmad Amine*, Lee Milburn*, Rajnish Gupta, Urara Kono, Rahul Mangharam

Penn Engineering

Traversing varying terrain presents challenges

Understanding vehicle-terrain interactions is critical to performant off-road control

Anticipating and reacting to terrain shifts, which are often abrupt, is critical to maintaining control

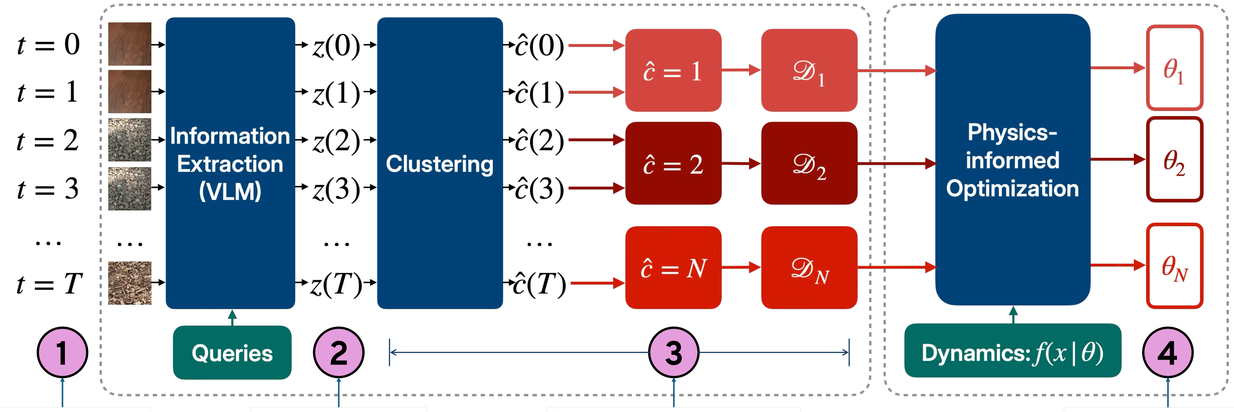
How can we traverse unknown numbers of unseen terrain, while remaining close to my reference trajectory?

1/5th-scale RoboRacer Autonomous Racecar

Our scale platform is based on a 1/5th scale Traxxas chassis, with compute provided by a Jetson AGX Orin. Images are provided by a GoPro Hero 11 Black, and state information is provided by a FixPosition Vision-RTK2

This enables:
- CLIP Inference for 30-60 images @ 1Hz
- MPPI control @ 120Hz
- Clustering at 1/15 Hz
- Physics-informed Optimization @ 1/10Hz

Our approach enables terrain identification with no prior information



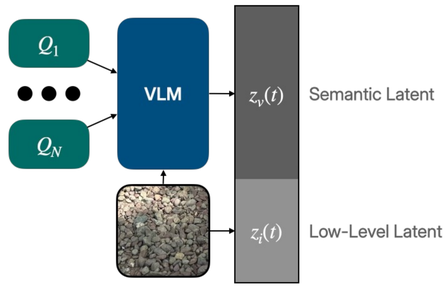
Observations are collected at each timestep, including **images, states, and controls**.

A **VLM**, relating queries to images, is used to **produce latent vectors** from the image.

These latent features are **clustered**. Data from timesteps with the same cluster are **aggregated into a dataset**.

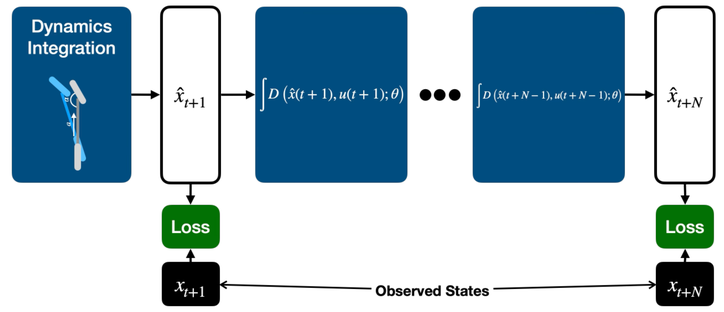
A gradient-based optimization algorithm is used to **identify parameters** for each cluster

Information Extraction



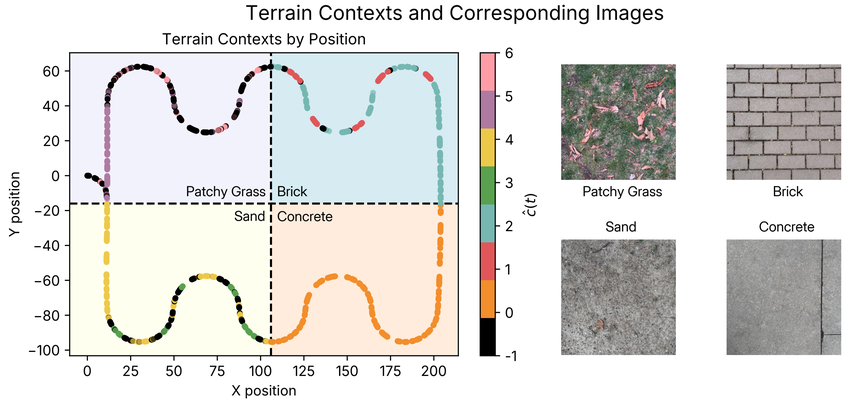
Images are **semantically compared to captions** in order to generate a low-dimensional semantic latent variable. This is concatenated to human-low level latents, like brightness or average color.

Physics-Informed Optimization

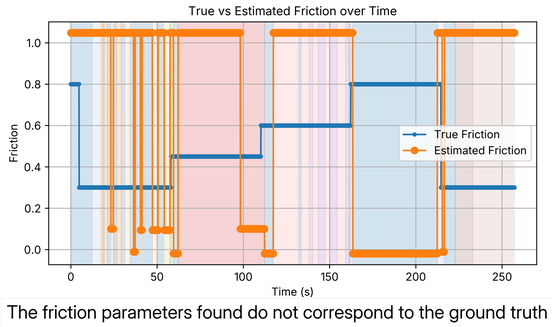


The physics-informed optimization procedure uses observed control inputs and states, combined with a given dynamics model to evaluate a set of parameters. Losses are backpropagated **onto the parameters** and optimized with an off-the-shelf optimizer.

Our approach enables zero-shot terrain identification



The figure shows the discovered terrain contexts in simulation experiments. The yellow, orange, light blue, and light purple correspond to patchy grass, dry soil, concrete, and brick respectively. The dots indicate the position and cluster. Black is used to show noise clusters, where no terrain context is identified. The clusters above were categorized in a single lap of the track, without any prior information about the environment. The textures used in the simulation are shown on the right hand side.



The friction parameters found do not correspond to the ground truth parameters, but do produce good control outcomes, avoiding catastrophic failure or control failure in simulated experiments

Ongoing Work

Renukanandan Tumu*, Ahmad Amine*, Lee Milburn, Antonio Loquercio, Rahul Mangharam

1. Replacing the clustering approach with a functional approximation to improve runtime
2. Identifying all parameters instead of only friction parameters