# Minimizing Expensive ML Operations in Video Data Exploration Tasks using Geo-Spatial Metadata

EPIC DATA lab UC Berkeley

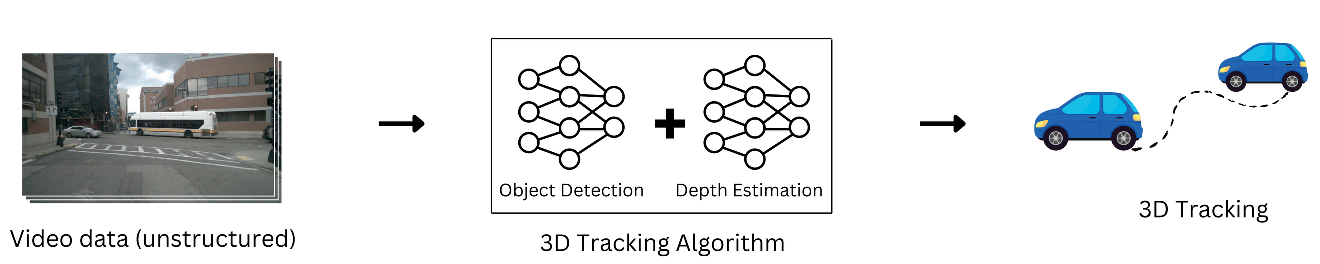
Chanwut (Mick) Kittivorawong | Yongming Ge | Yousef Helal | Daniel Kang | Alvin Cheung

## Background

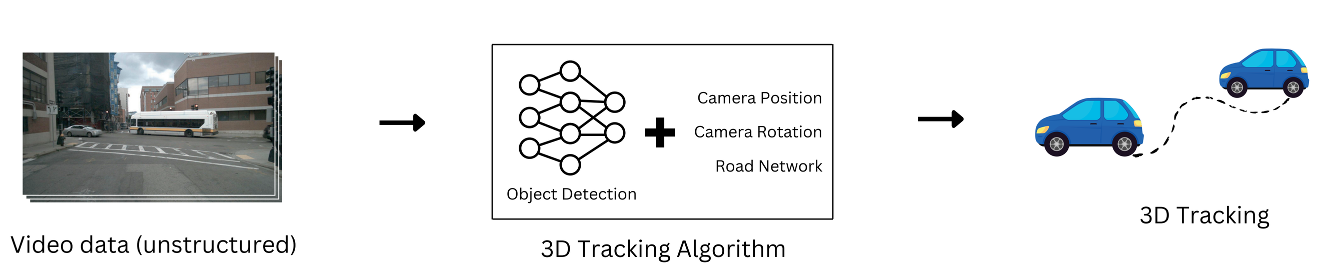
Video Data are unstructured. Analyzing them usually starts with using ML-based algorithms to extract all the necessary structured information from the videos. Then, we can search through these extracted information for objects/scenarios of interest. However, ML-based algorithms are typically slow. This project aims to find an alternative workflow that minimize the use of ML-based algorithms.

## Key Concepts

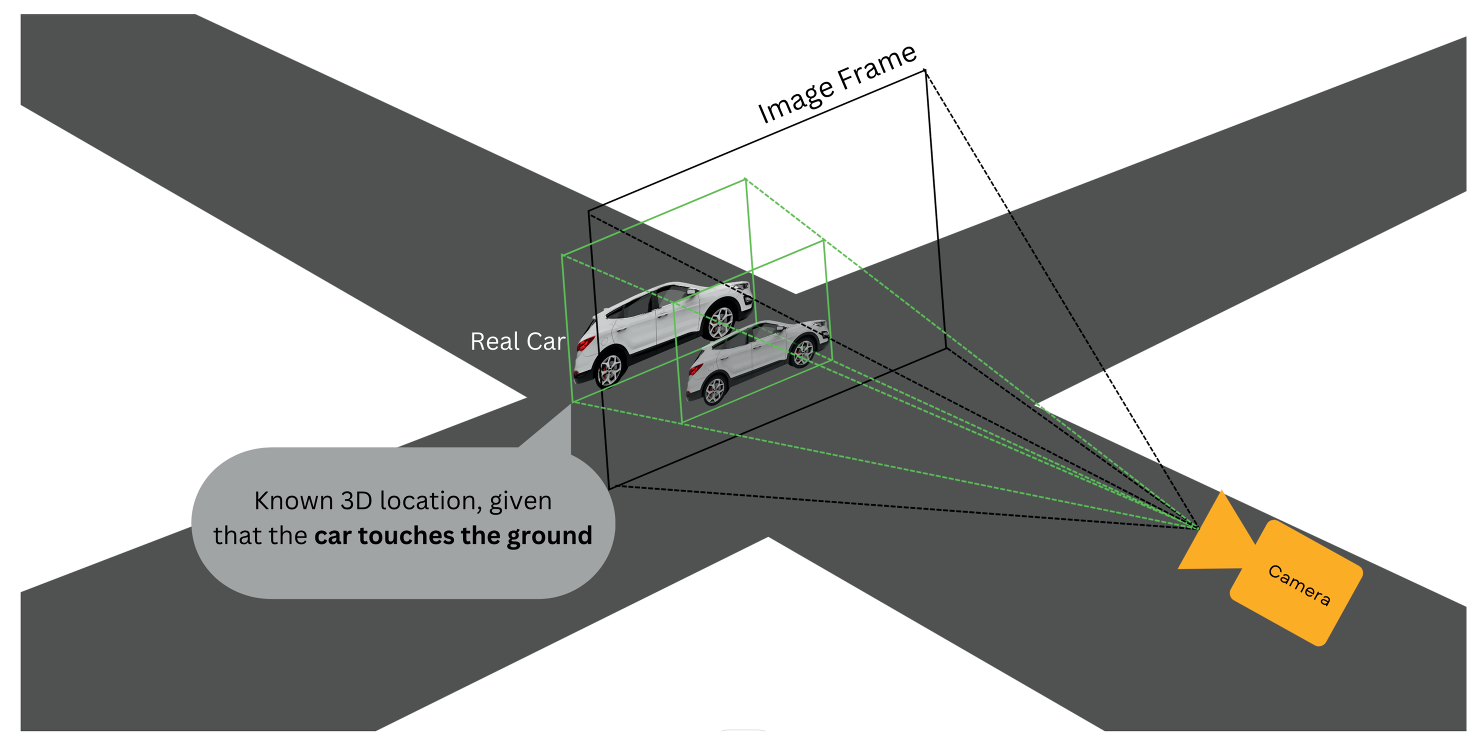
In this project, we focus on minimizing the use of ML-based algorithms for **3D-tracking** objects in videos.




Traditionally, we can use 2D-Object-Tracking algorithms to track objects. Then, we can use Mono Depth Estimation algorithm to estimate distance of the tracked objects from the camera, constructing 3D-tracking.

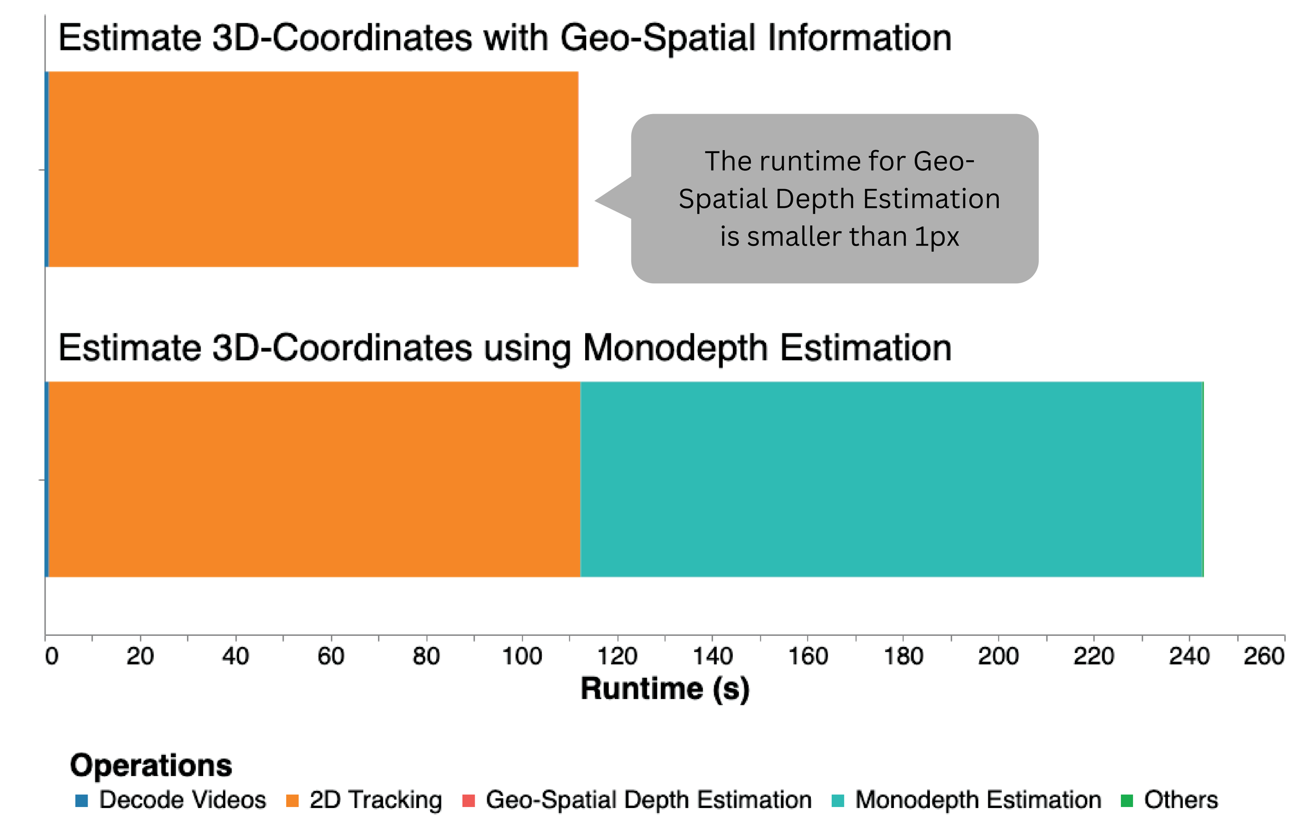



In this project, we introduce geo-spatial metadata, such as camera locations, camera rotations, and polygons representing road networks. We can use 2D-Object-Tracking algorithms to track objects. Then, we can estimate the 3D-location of each tracked object with the introduced geo-spatial metadata.



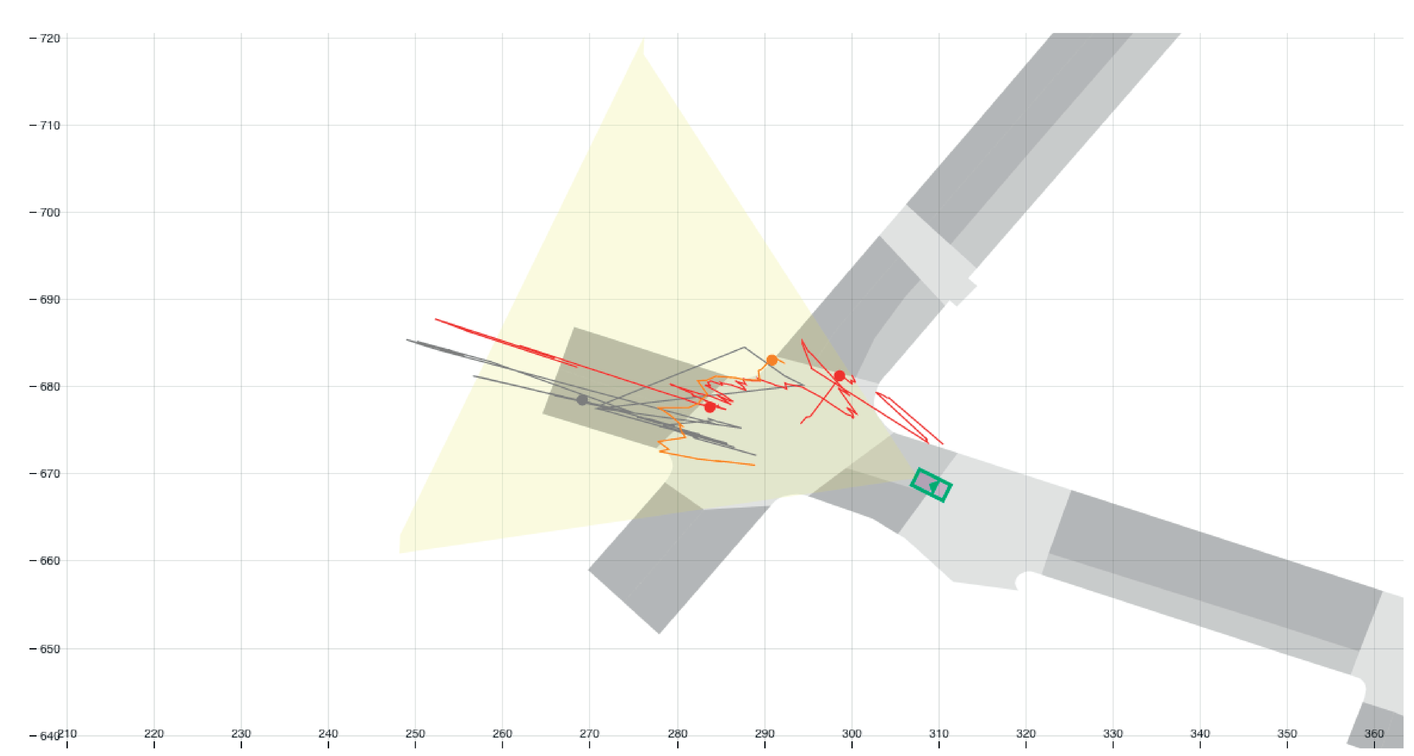

## 3D-Object Detection

In this project, we assume that ***"vehicles/persons always touch the ground"***. After we 2D-detect vehicles/persons, we can draw a line from the camera through the part of the detections that we know to touch the ground. We then find the intersection point of each line to ground in the 3D-world to get the estimated 3D-location of each detection.

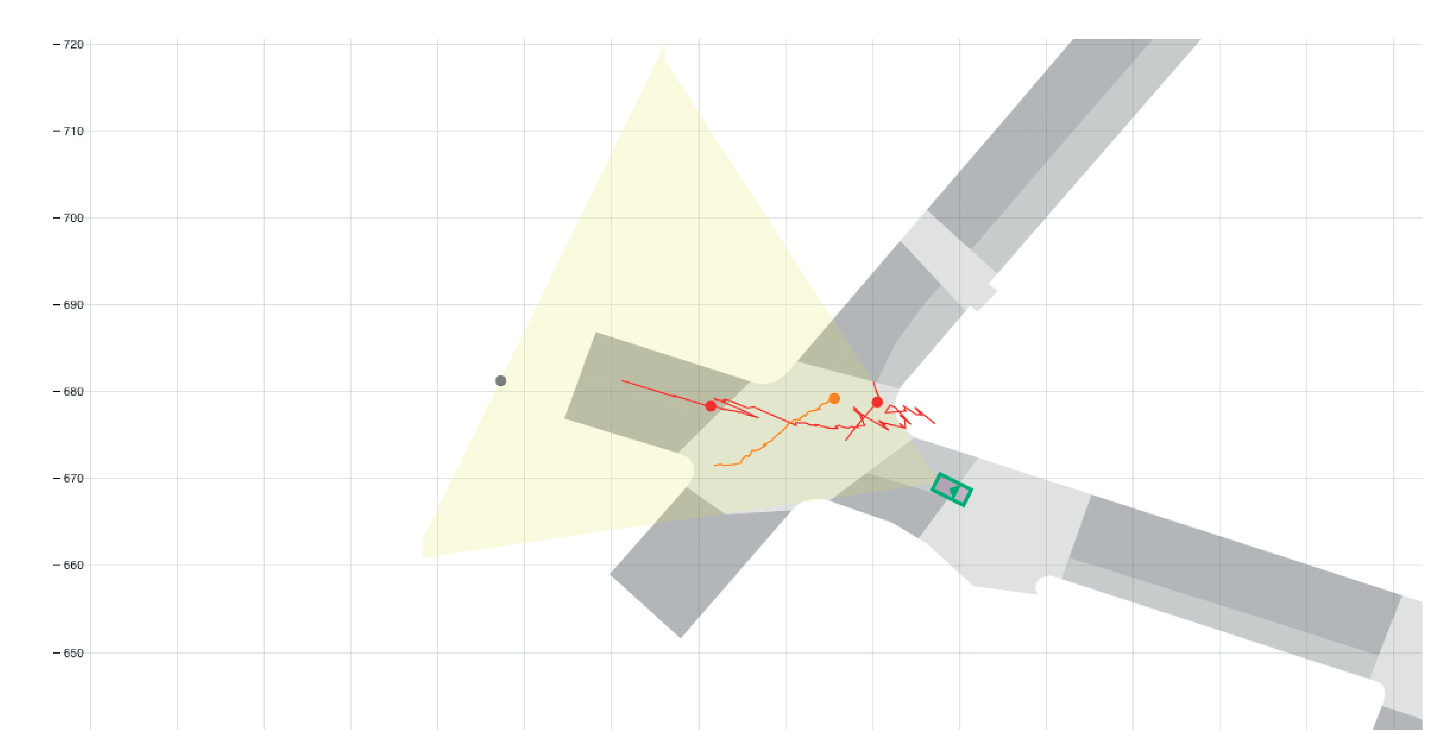



## Runtime Performance Comparison

Comparing the runtime of our 3D-tracking algorithm against the traditional ML-based algorithm (2D-tracking + mono depth estimation). Our depth estimation has an insignificant runtime compared mono depth estimation. Our algorithm reduces the runtime for 3D-tracking by half.



**Mono Depth Estimation**



**Geo-Spatial Depth Estimation**

We compare the 3D-tracking results using <u>mono depth estimation</u> and <u>geo-spatial depth estimation</u>. The green rectangle represents the camera, and the faded yellow triangle represents its view. Each line represents a tracking of each vehicle/person. Each dot represents a detection instance of a vehicle/person at a particular timestamp. Our algorithm results in smoother trackings comparing to the algorithm with the mono depth estimation.