

GROUNDDED: The Localizing Ground Penetrating Radar Evaluation Dataset

Teddy Ort, Igor Gilitschenski, and Daniela Rus
Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Email: {teddy, igilitschenski, rus}@mit.edu

Abstract—Mapping and localization using surface features is prone to failure due to environment changes such as inclement weather. Recently, Localizing Ground Penetrating Radar (LGPR) has been proposed as an alternative means of localizing using underground features that are stable over time and less affected by surface conditions. However, due to the lack of commercially available LGPR sensors, the wider research community has been largely unable to replicate this work or build new and innovative solutions. We present GROUNDDED an open dataset of LGPR scans collected in a variety of environments and weather conditions. By labeling this data with ground truth localization from an RTK-GPS / Inertial Navigation System, and carefully calibrating and time synchronizing the radar scans with ground truth positions, camera imagery, and Lidar data, we enable researchers to build novel localization solutions that are resilient to changing surface conditions. We include 108 individual runs totalling 450 km of driving with LGPR, GPS, Odometry, Camera, and Lidar measurements. We also present two new evaluation benchmarks for 1) Localizing in Weather and 2) Multi-lane Mapping, to enable comparisons of future work supported by the dataset. The dataset can be accessed at <http://lgprdata.com>.

I. INTRODUCTION

The ability to localize in the environment is critical to enable the widespread deployment of autonomous vehicles. While the Global Positioning System (GPS) is often used to obtain approximate global localization, it lacks the accuracy necessary to meet the stringent requirements of autonomous driving [1]. For this reason, most fielded autonomous vehicle solutions currently localize on HD maps with either lidar sensors [2]–[4], cameras [5], or both [6]. Localization with these sensors can provide accuracy in the range needed for autonomous operation. However, sensors that rely heavily on surface features in the environment have an inherent failure mode should the environment change between the mapping and localization phases. Some approaches aim to filter out dynamic objects during mapping [7]. Others seek to identify and map only stable features or landmarks in the environment [8], [9]. Robustly dealing with inclement weather such as snow is particularly challenging as snowfall can dramatically alter the surface appearance. Solving this problem remains one of the open challenges to enable human-level (or above) performance of autonomous vehicles in diverse environments.

Recently, Localizing Ground Penetrating Radar (LGPR) [10] has been proposed to address the localization task in such environments. By mapping and localizing using features beneath the ground, LGPR can avoid the instability of

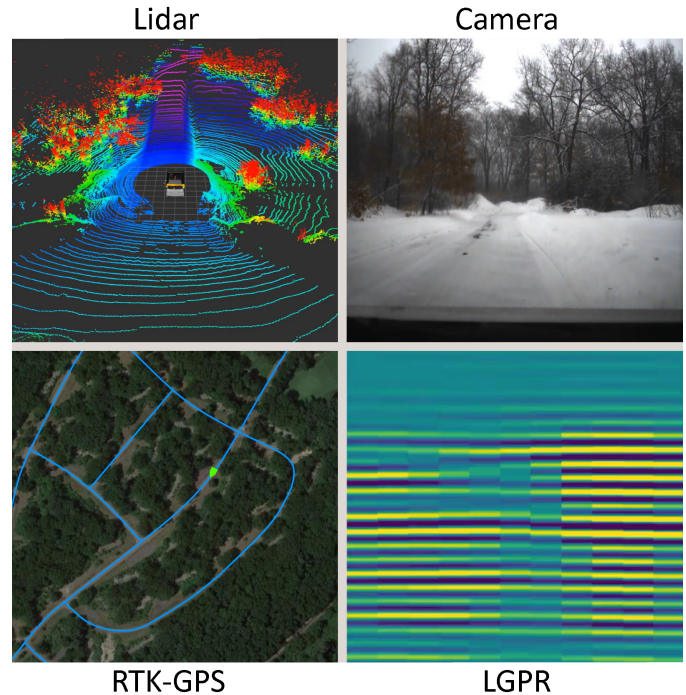


Fig. 1. The GROUNDDED dataset includes four data streams for each run. 1) Lidar scans from a roof mounted Velodyne HDL-64, 2) Camera images from a front-facing Point Grey Grasshopper camera, 3) Base station corrected RTK-GPS for groundtruth, and 4) LGPR data stream from the rear-mounted radar unit.

surface-based maps. LGPR was used as the sole localization sensor in [11] to navigate an automated vehicle in clear weather, rain, and snow conditions. However, the sensor demonstrated in that work is not commercially available. Typical Ground Penetrating Radar (GPR) sensors utilize frequencies in the 0.5-2.5 GHz range [12] to obtain a high-resolution image of objects close to the surface. However, a lower frequency of 100-400 MHz is ideal for localization because the greater penetration depth enables mapping of deeper, more stable features [10]. Several companies have recently announced plans to commercialize LGPR technology [13], [14]. Currently however, there are still no options to purchase LGPR sensors suitable for localization. The lack of publicly available sensors or datasets has left the technical community in a holding pattern.

In this work, we aim to address the lack of access to LGPR systems and to enable algorithmic development for localization of autonomous vehicles in a wide range of weather conditions and illumination. We describe and release an open dataset of LGPR frames collected using one of the prototype sensors described in [10]. We also propose two challenges aimed to accelerate the development of solutions for mapping and localization under challenging driving conditions such as difficult illumination, heavy rain, and snow. We believe the dataset will enable the research community to replicate and improve upon the current state-of-the-art, and to tackle new open problems for autonomous driving in difficult weather. For example, [11] found a degradation in localization performance when localizing in rain or snow, perhaps due to the unmodeled changes to the moisture content of the soil. Furthermore, both [10] and [11] conducted both mapping and localization in a single lane. Since practical autonomous vehicles will need to change lanes, it is an important area of research to stitch together multiple lanes to form a coherent map.

The dataset consists of 108 runs amounting to a total of 450 km and 12 hours of driving. For each route, there is data associated with clear, rainy, and snowy weather (see Fig. 2). The data includes groundtruth GPS location, odometry, and scans from the LGPR sensor, the camera, and the lidar system on the vehicle. Because the LGPR sensor has a data collection footprint equal to its width, which is smaller than the width of a road, a single path of the LGPR does not provide a complete map of the ground features for that road. We address this limitation by providing multiple paths for each segment, with the vehicle driving left, center, and right on the road (see Fig. 3), along with a challenge to align and stitch different paths into a complete map.

In summary, this paper contributes the following:

- The first publicly available dataset of Ground Penetrating Radar data for localization and mapping collected in a variety of weather conditions, and multiple adjacent lanes,
- Two challenge benchmarks 1) Localization in Weather and 2) Multi-lane Mapping to compare LGPR research
- Additional lidar and camera data streams to enable comparison with existing visual and lidar navigation approaches in driving.

The remainder of this paper is organized as follows: In the next section we review related work with LGPR sensors and autonomous driving datasets. In Sec. III we present the challenge benchmarks. Next, in Sec. IV we describe the dataset organization and software development kit. Then, in Sec. V we describe the research platform, sensor suite, calibration, and time synchronization of the sensor streams. Finally, in Sec. VI we conclude with some final thoughts regarding promising directions for future research.

II. RELATED WORK

Localization: Over the last two decades, the robotics community extensively considered the problem of localization and mapping [15] involving a diverse set of sensors, most

prominently cameras [16] and lidars [2]–[4]. Particularly for autonomous driving, a considerable amount of work focused on dealing with challenging and changing appearance conditions [9] such as weather [17] or occlusions [7], [18]. To improve robustness, radar has also been considered as a localization modality for autonomous driving [19]–[22]. Even with this additional modality, robust localization remains challenging due to phenomena such as occlusions. The goal of this dataset is to overcome or completely avoid some of these challenges. It enables wider research on a complimentary localization modality which does not suffer from occlusion by dynamic objects and changes in appearance conditions.

Ground Penetrating Radars: Only a few works considered the use of ground penetrating radars in robotics such as for landmine detection [23] or for autonomous surveys [24]. Using GPRs for localization has so far been considered only in [10], [11]. Consequently, most GPR datasets are targeted at very different application domains, e.g., for research on soil structure characterization [25] or meteorology [26]. There is currently no dataset allowing for wide-spread localization research with GPRs. The high cost of GPRs and mere prototype availability of GPRs specifically designed for localization makes research in this field for many groups completely impossible. By making their data publicly available, the authors aim to overcome this limitation simplifying research on radiogeological navigation.

Datasets: Because of the high cost of a retrofitted autonomous vehicle and to compare results more equally, a lot of Autonomous Driving research is already driven by benchmark datasets. In that context, the KITTI [27] dataset is one of the earliest and most popular in autonomous driving research. In recent years, numerous institutions made the data from their research vehicles publicly available [28]–[32] some of which also involve radar data [33], [34]. A dataset specifically focusing on radar perception is presented in [35]. More recently, several datasets covering novel sensing modalities have been made publicly available focusing on acoustic detection [36] and dynamic vision sensors (DVS) [37], [38]. In the same spirit, our work contributes ground penetrating radars as a new sensing modality to dataset-driven perception research.

III. BENCHMARK CHALLENGES

Prior localization results with LGPR have looked promising, yet there are two important limitations that must be overcome before LGPR sensors can be practically useful. The first requirement is to devise algorithms that can localize even when the prior map was recorded in different weather conditions. This can be challenging because LGPR data can be affected by the moisture content and temperature of the underground soil which can vary with surface weather conditions. In [11] a degradation in localization performance in rain and snow was measured, but their algorithm did not explicitly account for weather changes. The second requirement, is to build maps that can localize a vehicle while it is changing between multiple lanes. Since the LGPR sensor only records data directly beneath it, and the sensor only spans the width of the vehicle,

Year	Dataset	Location	Weather	Camera	Lidar	Radar	Other
2013	KITTI [27]	Karlsruhe	dry	✓	✓	-	
2016	Cityscapes [28]	50 cities	dry	✓	-	-	
2017	DDD17 [37]	Switzerland & Germany	dry, rain	✓	-	-	DVS
2017	Oxford [30], [33]	Oxford	dry, rain, snow	✓	✓	✓	
2018	ApolloScape [29]	4 x China	dry, rain, snow	✓	-	-	
2018	MVSEC [38]	Philadelphia	-	✓	✓	-	DVS
2019	Astyx [35]	-	-	✓	✓	✓	
2020	CADC [31]	Waterloo	dry, snow	✓	✓	-	
2020	nuScenes [34]	Boston, Singapore	dry, rain	✓	✓	✓	
2020	Waymo [32]	San Francisco, Phoenix, Detroit, Seattle, LA, Mountain View	dry, rain, snow	✓	✓	-	
2021	Delft [36]	Delft	-	✓	-	-	Mic. array
2021	Ours	Massachusetts	dry, snow, rain	✓	✓	LGPR	

TABLE I
OVERVIEW OF EXTEROCEPTIVE SENSING MODALITIES IN AUTONOMOUS VEHICLE NAVIGATION RESEARCH DATASETS.

prior work [10], [11] only used maps consisting of a single lane. Since practical autonomous vehicles will need the ability to maintain a seamless localization as they traverse multiple lanes, it will be necessary to devise mapping algorithms that can stitch together data from multiple passes in different lanes to obtain a cohesive road map. To ensure solutions to these limitations can be compared on an equal footing, we propose the following two challenge benchmarks. Our dataset specifically includes data to address these challenges including data collected in a variety of weather conditions and in multiple lanes as shown in Table III.

A. Localization in Weather Challenge

1) *Mapping and Localization Runs*: In the provided dataset, every run was collected as a pair to enable mapping and localization using the same environmental conditions. For example, *run_0001* and *run_0002* were both collected driving the same route, in an urban environment, in clear weather and in the right lane. All of this information can be found in the *runs.csv* file as described in Sec. IV-A. However, for the Localization in Weather Challenge, we aim to evaluate localization using maps that were created in different weather to demonstrate weather resilience. Therefore, we could instead evaluate localization using *run_0037* or *run_0038* which were both collected along the same route, but in snowy weather while still using *run_0001* or *run_0002* to build the map. In short, each run in the dataset includes a weather condition label [*clear*, *rain*, *snow*]. For this challenge, the mapping and localization runs should be along the same geographical route, but in different weather conditions.

In the event learning-based algorithms are utilized for mapping, we would like to clarify here that it is acceptable to include runs which were collected in inclement weather in the training phase, as long as they were collected at a different location. This mirrors the real-world constraints where one could envision training an algorithm to localize to single weather maps in all weather by including a small sample

of varying weather in the training phase. Finally, while other onboard sensors (e.g. odometry, camera, or lidar) may be used to enhance the localization estimate, the estimate at each time t_i may only use measurements taken at a prior time t_j such that $j \leq i$.



Fig. 2. Left: Camera images from three different runs of the same route (route_04) capturing LGPR data in the same location in clear weather, snow, and rain to support the Localization in Weather Challenge. Right: Trajectory of route_04 overlaid on a map. Other runs capture a variety of environments including rural (shown here), urban, and highway.

2) *Localization in Weather Metric*: The metric used for localization in weather is an Absolute Trajectory Evaluation. Since the LGPR data is labeled with ground truth RTK-GPS data, we can directly evaluate the trajectory error rather than implementing a relative metric as is often needed with SLAM or VIO solutions [39]. Additionally, we limit the benchmark

to the two translation degrees of freedom and single rotation (yaw) on the surface of the ground as these are the critical values necessary for autonomous driving. Thus, the first two metrics are the Root Mean Square Error (RMSE) of the translation and orientation of the vehicle t_{rmse}, θ_{rmse} evaluated over the entire run. Next, because for driving purposes, the lateral translation error (with respect to the lane) is often far more important than the longitudinal, we further decompose the error into its lateral and longitudinal components to obtain two additional error metrics: t_{lat}, t_{long} . While all four of these error metrics should be reported, for the purpose of providing an overall score we propose a weighting:

$$s = t_{lat} + 0.1t_{long} + 10\theta_{rmse} \quad (1)$$

Intuitively, this assigns equal cost to 10 cm of error in the lateral lane direction, 1 m of error in the longitudinal lane direction, and 0.57° of orientation error which have similar real-world importance.

B. Multi-lane Mapping Challenge

1) *Mapping and Localization Runs*: In the provided dataset, each run is labeled with a lane attribute in [Left, Center, Right, Changing] as shown in Fig. 3. These lanes are overlapping since the center lane is not a separate lane, but rather the sensor centerline is passing over the lane divider and including partial data from each of the left and right lanes. The purpose of these runs is specifically to support building cohesive maps that can track a vehicle even while it is changing lanes, or only partially overlapping a lane. For this challenge, multiple runs where the lane is in [Left, Center, Right] in the same route should be used for map creation. In runs where the lane is marked Changing, the vehicle was driven along the route while randomly choosing any of the [Left, Center, Right] lane positions and changing between them. Those runs should only be used for localization evaluation and not included in mapping or training data.

2) *Multi-lane Mapping Metric*: The metric used for the Multi-lane Mapping benchmark is similar to that described for the previous benchmark. One important difference is that here we explicitly do not follow a single lane. Therefore, lateral and longitudinal lane errors are not needed and only the absolute trajectory errors t_{rmse}, θ_{rmse} are necessary. The overall score is then calculated as:

$$s = t_{rmse} + 10\theta_{rmse} \quad (2)$$

IV. DATASET

The dataset is available for download at <http://lgprdata.com>. The data is stored hierarchically as shown in Fig. 4.

A. Run Level Data

At the highest level are runs. Each run is a single data collection instance often taking the form of a loop. To avoid potential issues with overlapping sensor data, the start and end points of the loop always have a small gap of ≈ 10 m. For each run, we provide run level information in a file called

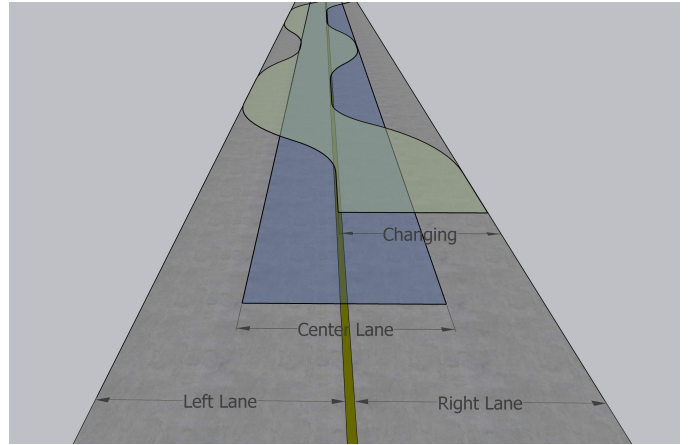


Fig. 3. The Multi-lane Mapping Challenge provides LGPR frames collected in four lane positions: 1) Left, 2) Right, and 3) Center 4) Changing. These runs can be used to create a consistent map capable of localizing the vehicle continuously even while switching lanes.

runs.csv. This file contains columns as shown in Table II. Each row in *runs.csv* has a unique *run_id* which corresponds to a directory such as *run_0001* in the runs directory (see Fig. 4). Note that the *route_id* corresponds to a unique route traversal but duplicates do exist because the same route was driven in multiple runs. For example, *run_0001* and *run_0002* represent two unique runs, but have the same *route_id* because they traversed the same physical route. Furthermore, a single route can be traversed in any of the lanes, or in either direction. One final caveat to bear in mind is that the direction of traversal rotates the semantic meaning of the lane column. Thus, two runs with the same *route_id* but different direction values would imply traversing the same lane in opposite directions if the lane value in one was equal to *right* while in the other was *left*. In general, the purpose of grouping runs by route, lane, and direction, is to provide data for the Multi-lane Mapping Challenge (Sec. III-B). For the simple case of mapping and localizing on the same trajectory, (as in the prior work), one can simply use runs with identical route, lane, and direction.

Column	Data Type	Description
<i>run_id</i>	Integer	A unique value for each run in the dataset
<i>date</i>	Date	The date when the run was recorded
<i>road_type</i>	String	One of {urban, rural, highway}
<i>route_id</i>	Integer	A unique value for each route
<i>weather</i>	String	One of {clear, rain, snow}
<i>direction</i>	Integer	{-1, 1} = {clockwise, anticlockwise}
<i>lane</i>	String	One of {left, center, right, changing}
<i>length</i>	Float	The total length (km) of the run
<i>duration</i>	Float	The total duration (s) of the run
<i>sensors</i>	List	A list of comma separated sensor names

TABLE II

THE FILE *runs.csv* CONTAINS METADATA FOR EVERY RUN IN THE DATASET ORGANIZED INTO THE COLUMNS SHOWN.

The dataset contains runs in a variety of lanes, and environmental and weather conditions to support the benchmark challenges proposed earlier. Table III shows the splits in the data for each of the relevant data conditions.

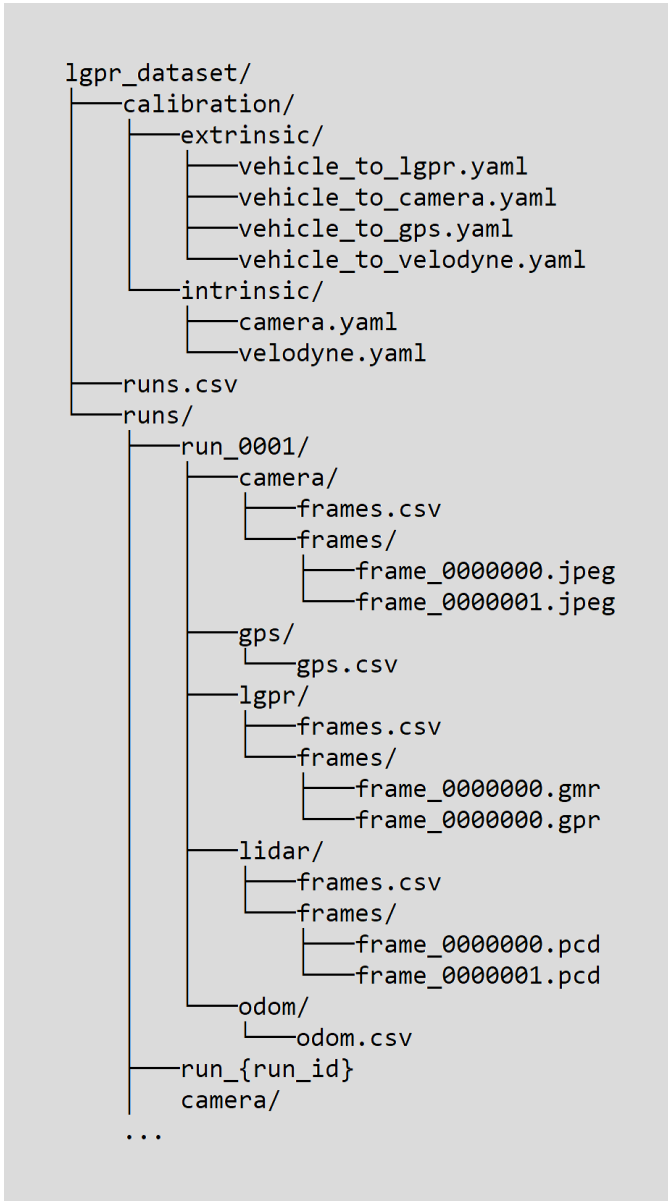


Fig. 4. The dataset is organized in directories for each run. Run metadata such as the lane and weather conditions is provided in the `runs.csv` file and referenced by the corresponding `{run_id}`. Similarly, frame metadata is provided in a `frames.csv` file referencing individual frames by their `{frame_id}`.

B. Sensor Level Data

Within each run directory, there are several directories, one for each element in the sensors field in the corresponding row of `runs.csv`. Currently, every run includes at a minimum [`lgpr`, `gps`, `odom`]. Many runs additionally include [`camera`, `lidar`]. Next, we briefly describe the data formats of each of these sensor streams, for additional details, please see the dataset documentation.

1) `lgpr`: The LGPR sensor frames are arrays with shape (11, 369) corresponding to the 11 radar channels (11 Tx and Rx pairs for the 12 radar elements in the array) and the

369 depth bins. Each value in the array is an `int8` ranging from [-128, 127]. The `lgpr` directory contains a file `frames.csv` which includes a table containing the `frame_id` for each LGPR frame in the run, along with a timestamp providing the synchronized time (see Sec. V-C) for that frame. In the `frames` directory is included two files for each frame in `frames.csv`: 1) `frame_id.gpr` contains the raw LGPR array in csv format while 2) `frame_id.gmr` contains the processed LGPR scan with mean removed as described in Sec. V-D. A script for reading and visualizing these scans is provided in the SDK.

2) `gps`: The ground truth GPS data includes position and velocity. Each GPS reading contains sixteen values including [`longitude`, `latitude`, `altitude`], position and orientation in the UTM frame [`x`, `y`, `z`, `qx`, `qy`, `qz`, `qw`], and velocity in the vehicle frame [`\dot{x}` , `\dot{y}` , `\dot{z}` , `\dot{r}` , `\dot{p}` , `\dot{y}`]. Note that most GPS measurements include Differential GPS (DGPS) corrections received from a fixed base station. However, because our base station has a range limited to ≈ 10 km, runs with `road_type` equal to `highway` in the `runs.csv` file do not have base station corrections. All of these twelve columns are included, along with a timestamp in a file called `gps.csv` which includes all GPS measurements for the run.

3) `odom`: The odometry data includes measurements from the wheel encoders on each of the rear wheels and the Inertial Measurement Unit (IMU). For each wheel encoder measurement we include the total distance in meters, traversed by the left and right rear wheels respectively along with a timestamp. These measurements are provided in `encoder.csv`. Next, the IMU data is a vector of ten values including 3-vectors for linear acceleration and angular velocity, and an orientation quaternion. These values, along with a timestamp, are provided for each IMU measurement. Finally, for convenience, we also provide `odom.csv` which contains the vehicle position [`x`, `y`, `z`] and orientation quaternion [`qx`, `qy`, `qz`, `qw`] with respect to the start frame calculated by accumulating the wheel encoder and IMU measurements using a Kalman filter.

4) `camera`: The LGPR sensor is particularly useful in situations where camera sensors can fail such as glare, darkness, or inclement weather. For this reason, we include camera data for comparison purposes. We use a PointGrey Grasshopper front-facing camera mounted behind the windshield. Each frame has resolution 1928x1448. The file `camera.csv` contains a table with columns for the `frame_id` of each camera frame along with the timestamp. The directory `frames/` contains camera images of the form `frame_id.png` where each image corresponds to the `frame_id` in the `frames.csv` file.

5) `lidar`: Similar to the camera sensor, we provide lidar data primarily as a baseline comparison tool. The lidar sensor is a Velodyne HDL-64 mounted above the roof of the vehicle offset toward the front (see Fig. 5). It provides pointcloud data where each scan contains $\approx 191,000$ data points, and each point is a 4-vector of [`x`, `y`, `z`, `intensity`] giving the geometric position of the point with respect to the sensor frame, and the intensity of the laser reflection. We accumulate measurements until a 360° rotation is completed at a rate of 10 Hz and include one scan/revolution. The file `lidar.csv` contains a table with a col-

	Type	Total Length (km)
Road Type	highway	316.2
	rural	115.2
	urban	17.6
Weather	clear	151.8
	rain	145.4
	snow	151.8
Lane	center	21.6
	changing	16.8
	left	38.4
	right	372.2

TABLE III

THE GROUNDED DATASET CONTAINS DATA COLLECTED IN A VARIETY OF LANE POSITIONS, AND ENVIRONMENT AND WEATHER CONDITIONS TO SUPPORT THE PROPOSED BENCHMARK CHALLENGES.

umn *frame_id* along with a timestamp. The directory *frames/* contains files *frame_id.pcd* which includes the pointclouds in PCD format [40].

V. DATA COLLECTION PLATFORM

A. Vehicle Infrastructure

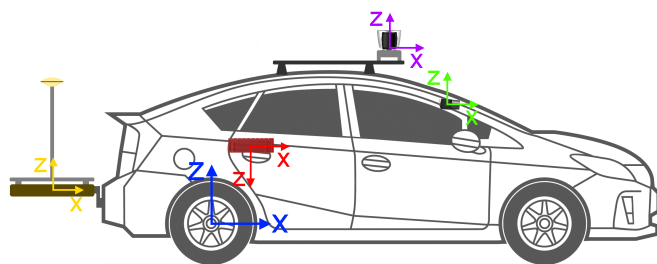


Fig. 5. Top: The Toyota Prius vehicle platform used to collect the dataset. Bottom: A schematic showing the positions of reference frames for the vehicle and each of the sensors. Blue is the vehicle frame, Red is the OXTS GPS/IMU, Green is the camera, Magenta is the lidar, and Yellow is the LGPR sensor. The measured transforms between these frames are included in the dataset.

The dataset was collected using a Toyota Prius research platform shown in Fig. 5. The sensors mounted on the vehicle include wheel encoders, an IMU, GPS, lidar, camera, and the LGPR radar array. The origin frame for the vehicle is located in the center of the rear axle, with the x -axis parallel to the ground and oriented toward the front of the vehicle, and the z -axis oriented upward. Fig. 5 shows the reference frames for all of the other sensors as well. Note that all of the sensors are located on the central plane of the vehicle (XZ -plane) except

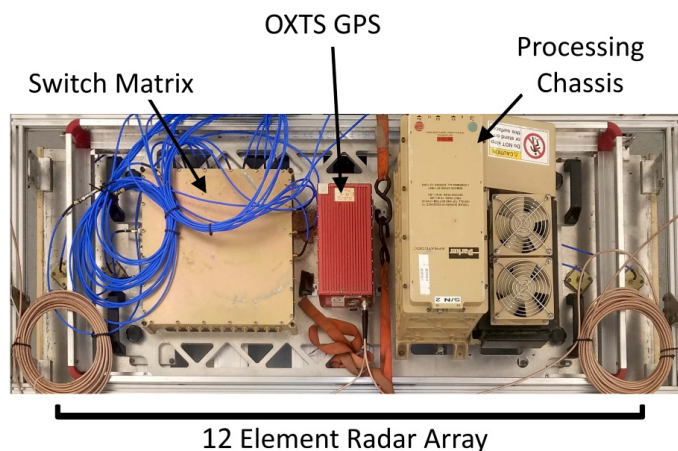


Fig. 6. The LGPR sensor used to collect the dataset. The processing chassis communicated with the switch matrix to control transmit and receive on the 12 radar elements. The GPS shown here was used for time synchronization, but groundtruth information was obtained by a separate GPS unit onboard the vehicle.

for the GPS unit which is slightly offset. The actual values for all of these transforms are provided in the *calibration/extrinsic* directory as shown in Fig. 4. In the following subsections, we describe the details of each sensor and how the data was collected, synchronized, and calibrated.

The data was collected on two computers. The first one, *CAR_PC* was connected to a vehicle-wide LAN which received data from the GPS, IMU, wheel-encoders, and lidar. It also utilized a USB hub to receive images from the camera. Due to technical limitations of the LGPR sensor prototype, it could not be configured to stream the raw data to the vehicle LAN in real time. Instead, a second computer, *LGPR_PC* was used to receive and store the LGPR frames separately. In Sec. V-C we describe how we utilize an accurate GPS time reference to provide time-synchronized data streams. The *CAR_PC* is a Dell Laptop with an $i7$ processor running Ubuntu 18.04. The *LGPR_PC* is a Single Board Computer (SBC) onboard the LGPR sensor prototype and accessed via Ethernet for data retrieval.

B. Sensors

1) *LGPR Sensor*: The LGPR sensor was mounted on the rear of the vehicle by attaching to the trailer hitch. It is a completely self-contained unit only connected to the vehicle for power. Here we describe the main components of the LGPR sensor, for more details see [10] which describes the design of the sensor. This sensor is not commercially available and there are only a few prototypes which is the primary motivation for sharing the data in this work.

As seen in Fig. 6 the main LGPR sensor components include a 12 element radar array, a switch matrix, an OXTS RTK-GPS unit, and a processing chassis. The radar array transmits on a single element at a time, while receiving on the next element. Each pair of elements thus provides a single channel of data which produces images 11 pixels wide. Note that the array is

152 cm wide (the same as the width of the vehicle) and 61 cm from front to back. While the GPS unit contained in the LGPR sensor is typically used for onboard mapping, here we use it only for time-synchronization, instead using the more accurate base-station corrected GPS measurements taken onboard the vehicle for groundtruth as described in Sec. V-B2

At runtime, the array completes a sweep of all 11 channels at 126 Hz and these data are binned into 369 depth bins to provide an 11x369 pixel image that spans the width of the vehicle and penetrates approximately 3 m beneath the ground (the actual depth can vary by region and soil content). The switch matrix receives these signals and transmits them to the processing chassis where the radar images are stored for later retrieval.

2) *RTK-GPS*: An OXTS RT3003 Inertial Navigation System (INS) was used to provide a groundtruth vehicle position for all runs. To achieve an accuracy of ≈ 2 cm, the GPS sensor requires corrections to the raw GPS signal from a fixed base station. Note that for all runs with *road_type* equal to *rural* or *urban* we placed and calibrated the base station in the test region as described in Sec. V-D. However, since the base station range is limited to 10 km, the runs with *road_type* equal to *highway* do not include base station corrections.

3) *Odometry*: The odometry sensors include two magnetic hall effect wheel encoders one in each of the rear wheels. These provide counts as each pole passes the sensor. Additionally, we utilize the IMU built in to the RT3003 to obtain acceleration and angular rates. Together, the wheel odometry and IMU data can be used to obtain an interoceptive estimate of the vehicle motion independent from the exteroceptive measurements of the GPS and LGPR sensors.

4) *Camera*: We utilize a front-facing Point Grey Grasshopper camera with a resolution of 1928x1448 at 6fps. The camera is mounted behind the windshield to protect it from rain or snow. Windshield wipers were utilized when required to ensure the windshield remained clear. The main motivation for providing camera imagery in this work is to serve as a baseline to compare how weather effects vision sensors compared to LGPR.

5) *Lidar*: A Velodyne HDL-64 sensor is mounted on the roof of the vehicle. We mount the sensor on the center plane of the vehicle, slightly forward to obtain a better view of the road in front of the vehicle at the expense of some occlusion caused by the vehicle itself in the rear. The Velodyne scanner spins at 600 RPM yielding 360-degree scans at 10 Hz. The scans are labeled with the synchronized time at the end of each revolution, to obtain the time of individual points, one can linearly interpolate between the time stamps for each azimuth.

C. Time Synchronization

As noted in Sec. V-A, the LGPR sensor data is collected in isolation from the other vehicle sensors. We utilize the GPS time reference on the *CAR_PC* to obtain a single reference that is accurate to within a few nanoseconds [41]. First, the GPS position for each LGPR frame is recorded based on the GPS device incorporated within the LGPR sensor. Next, the

GPS sensor within the vehicle records a data-stream of pairs of time stamps (t_{sys}, t_{gps}) at 100 Hz. Lastly, for each vehicle sensor (GPS, odometry, camera, and lidar), each data point is recorded with the corresponding system time s_{sys} .

Then, to synchronize all of the onboard sensor data in post-processing, we interpolate the GPS reference time for each sensor data point as:

$$s_{gps} = \frac{s_{sys} - t_{sys}^i}{t_{sys}^{i+1} - t_{sys}^i} (t_{gps}^{i+1} - t_{gps}^i) \quad (3)$$

where $(t_{sys}^i - t_{gps}^i)$ is the closest time reference pair with $t_{sys}^i < s_{sys}$ and $(t_{gps}^{i+1} - t_{gps}^i)$ is the next consecutive time reference pair.

The data stream provided by the LGPR sensor contains geo-referenced radar frames. However, since the position estimates are obtained with the standard quality GPS device integrated in the LGPR sensor, rather than the RTK-GPS system onboard the vehicle used for ground truth, we cannot simply match the LGPR frames to the ground truth positions using their recorded positions. Instead, to obtain accurate positions for LGPR frames, we first obtain timestamps by noting that the radar frames are collected at a fixed frequency of ≈ 126 Hz. Therefore, to obtain timestamps, the LGPR stream is aligned with the GPS data stream. We first differentiate the positions from both the LGPR scans and the RTK-GPS systems using a centered Savitzky-Golay filter [42]. Next, we obtain the alignment offset between the velocities by maximizing the Pearson correlation coefficient between the velocities using numerical optimization. The result of this maximization, combined with the fixed frequency of the LGPR scans allows every LGPR scan to be labeled with a corresponding timestamp on the GPS clock.

D. Calibration

Each of the sensors included in the data stream was calibrated as needed before each run. Here we describe the various calibration processes for each sensor. Additionally, when possible we include the calibration files in the dataset (see Fig. 4).

1) *LGPR Sensor Calibration*: The LGPR sensor needs to be calibrated to ensure changing environmental conditions do not unduly affect the sensor readings. The sensor includes a calibration routine which pulses each element for a short time and measures and records the mean values. This enables the intensity data to be stored in a mean-centered format which helps remove any biases due to ambient conditions. This allows the device to operate reliably in temperatures ranging from -5°C to 50°C [10]. We calibrated the LGPR sensor at the start of each day. Additionally, in the event the temperature changed drastically throughout the day, the calibration routine was repeated in between runs. For maximum flexibility, we include for each frame both the raw LGPR frame as *frame_id.gpr*, and the mean-centered frame as *frame_id.gmr*. The transform between the vehicle frame and LGPR frame is provided in *vehicle_to_lgpr.yaml* (see Fig. 4).

2) *RTK-GPS Calibration*: The RTK-GPS system requires calibration in two ways. Firstly, because the GPS is receiving base station corrections, all measurements are with respect to the fixed location of the base station. Any errors in the measurement of the location of the base station itself would propagate to the vehicle measurements and reduce the system accuracy. To mitigate this, we fix the location of the base station by mounting it in a permanent position. For each of the road types *rural* and *urban* we selected a fixed base station antenna mount and recorded and averaged GPS positions for one hour. Once that averaging period was complete, we record the mean position and use it for all future runs using that base station position. This ensures that no errors are introduced between runs due to incorrect measurement of the base station position. Note that for *road_type* highway we do not use corrections because of the limited base station range.

The second calibration necessary for the GPS groundtruth is a built-in calibration routine in which the vehicle is driven in several loops and ∞ -paths. The manufacturer provides a software tool for using these drives to fine-tune the position and orientation of the sensor within the vehicle, as well as the relative positions of each of the two roof-mounted GPS antennas. We store those values onboard the device, and provide them with the dataset in *vehicle_to_gps.yaml* (see Fig. 4).

3) *Camera Calibration*: To calibrate the Point Grey Grasshopper front-facing camera, we utilize the camera calibration method described here [43]. The intrinsic calibration file is included in the dataset as *intrinsic/camera.yaml*. The measured extrinsic calibration between the camera and the velodyne is provided in *velodyne_to_camera.yaml* as shown in Fig. 4.

4) *Velodyne Calibration*: The Velodyne sensor includes a factory provided calibration file that accounts for small differences in manufacturing and assembly and the affect they have on the conversion between the measured azimuth/angle of each point, and its position in the sensor frame. We apply this calibration file in order to produce the Velodyne frames found in *frame_id.pcd*. We also include the calibration file in *intrinsic/velodyne.xml*. Additionally, we provide the transform between the Velodyne sensor frame and the vehicle frame in *vehicle_to_velodyne.yaml* as shown in Fig. 4.

VI. CONCLUSION

In this work, we presented a novel dataset for localization and mapping research in autonomous driving using a ground penetrating radar. Our goal is to unlock the potential of radiogeological navigation as this sensing modality holds the promise to enable autonomous vehicle localization even in the most challenging weather conditions. Together with the dataset, this work proposes evaluation protocols and additionally provides camera and lidar data to simplify comparisons with established algorithms on these sensing modalities.

ACKNOWLEDGMENTS

This work was supported in part by the MIT Lincoln Laboratory and ONR grant N00014-18-1-2830. The authors

would also like to thank Mike Boulet and Byron Stanley for their continued advice and technical support throughout this project.

REFERENCES

- [1] M. G. Wing, A. Eklund, and L. D. Kellogg, "Consumer-grade global positioning system (GPS) accuracy and reliability," *Journal of forestry*, vol. 103, no. 4, pp. 169–173, 2005.
- [2] J. Levinson, M. Montemerlo, and S. Thrun, "Map-Based Precision Vehicle Localization in Urban Environments," in *RSS*, 2007.
- [3] J. Levinson and S. Thrun, "Robust Vehicle Localization in Urban Environments Using Probabilistic Maps," in *ICRA*, 2010.
- [4] R. W. Wolcott and R. M. Eustice, "Fast LIDAR localization using multiresolution Gaussian mixture maps," in *ICRA*, 2015.
- [5] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "ORB-SLAM: a versatile and accurate monocular SLAM system," *IEEE transactions on robotics*, vol. 31, no. 5, pp. 1147–1163, 2015.
- [6] R. W. Wolcott and R. M. Eustice, "Visual localization within lidar maps for automated urban driving," in *IROS*, 2014.
- [7] B. Bescos, J. M. Facil, J. Civera, and J. Neira, "DynaSLAM: Tracking, mapping, and inpainting in dynamic scenes," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 4076–4083, 2018.
- [8] M. Dymczyk, E. Stumm, J. Nieto, R. Siegwart, and I. Gilitschenski, "Will it last? Learning stable features for long-term visual localization," in *2016 Fourth International Conference on 3D Vision (3DV)*. IEEE, 2016, pp. 572–581.
- [9] M. Burki, C. Cadena, I. Gilitschenski, R. Siegwart, and J. Nieto, "Appearance-based landmark selection for visual localization," *Journal of Field Robotics*, vol. 36, no. 6, pp. 1041–1073, 2019.
- [10] M. Cornick, J. Koechling, B. Stanley, and B. Zhang, "Localizing Ground Penetrating RADAR: A Step Toward Robust Autonomous Ground Vehicle Localization," *JFR*, vol. 33, no. 1, pp. 82–102, 2016.
- [11] T. Ort, I. Gilitschenski, and D. Rus, "Autonomous Navigation in Inclement Weather Based on a Localizing Ground Penetrating Radar," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 3267–3274, 2020.
- [12] A. Benedetto, F. Tosti, L. B. Ciampoli, and F. D’amico, "An overview of ground-penetrating radar signal processing techniques for road inspections," *Signal processing*, vol. 132, pp. 201–209, 2017.
- [13] D. Ryan, "Lincoln Laboratory enters licensing agreement to produce its localizing ground-penetrating radar," <https://news.mit.edu/2017/lincoln-laboratory-enters-licensing-agreement-to-produce-localizing-ground-penetrating-radar-0718>, 2017.
- [14] K. Shaw, "Meet WaveSense: Providing Accuracy for Vehicles Via Underground Radar," <https://www.roboticsbusinessreview.com/unmanned/unmanned-ground/meet-wavesense-providing-accuracy-for-vehicles-via-underground-radar>, 2020.
- [15] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard, "Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age," *IEEE Transactions on Robotics*, vol. 32, no. 6, 2016.
- [16] S. Lowry, N. Sunderhauf, P. Newman, J. J. Leonard, D. Cox, P. Corke, and M. J. Milford, "Visual Place Recognition: A Survey," *IEEE Transactions on Robotics*, vol. 32, no. 1, 2016.
- [17] A.-D. Doan, Y. Latif, T.-J. Chin, Y. Liu, T.-T. Do, and I. Reid, "Scalable Place Recognition Under Appearance Change for Autonomous Driving," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.
- [18] M. Fehr, F. Furrer, I. Dryanovski, J. Sturm, I. Gilitschenski, R. Siegwart, and C. Cadena, "TSDF-based change detection for consistent long-term dense reconstruction and dynamic object discovery," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 2017.
- [19] K. Burnett, A. P. Schoellig, and T. D. Barfoot, "Do We Need to Compensate for Motion Distortion and Doppler Effects in Spinning Radar Navigation?" *IEEE Robotics and Automation Letters*, vol. 6, no. 2, 2021.
- [20] M. Rapp, M. Hahn, M. Thom, J. Dickmann, and K. Dietmayer, "Semi-Markov Process Based Localization Using Radar in Dynamic Environments," in *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, 2015.
- [21] K. Werber, J. Klappstein, J. Dickmann, and C. Waldschmidt, "Point group associations for radar-based vehicle self-localization," in *International Conference on Information Fusion (FUSION)*, 2016.

- [22] —, “Association of Straight Radar Landmarks for Vehicle Self-Localization,” in *Intelligent Vehicles Symposium (IV)*, 2019.
- [23] K. M. Dawson-Howe and T. G. Williams, “The detection of buried landmines using probing robots,” *RAS*, 1998.
- [24] R. M. Williams, L. E. Ray, and J. Lever, “An autonomous robotic platform for ground penetrating radar surveys,” in *IGARSS*, 2012.
- [25] A. Romero-Ruiz, N. Linde, T. Keller, and D. Or, “A Review of Geophysical Methods for Soil Structure Characterization,” *Reviews of Geophysics*, vol. 56, no. 4, pp. 672–697, 2018.
- [26] T. Kubota, T. Ushio, S. Shige, S. Kida, M. Kachi, and K. Okamoto, “Verification of High-Resolution Satellite-Based Rainfall Estimates around Japan Using a Gauge-Calibrated Ground-Radar Dataset,” *Journal of the Meteorological Society of Japan. Ser. II*, vol. 87A, 2009.
- [27] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, “Vision meets robotics: The KITTI dataset,” *The International Journal of Robotics Research*, vol. 32, no. 11, 2013.
- [28] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, “The Cityscapes Dataset for Semantic Urban Scene Understanding,” in *(CVPR)*, June 2016.
- [29] X. Huang, X. Cheng, Q. Geng, B. Cao, D. Zhou, P. Wang, Y. Lin, and R. Yang, “The ApolloScape Dataset for Autonomous Driving,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 954–960.
- [30] W. Maddern, G. Pascoe, C. Linegar, and P. Newman, “1 Year, 1000 Km: The Oxford RobotCar Dataset,” *The International Journal of Robotics Research*, vol. 36, no. 1, pp. 3–15, 2017.
- [31] M. Pitropov, D. E. Garcia, J. Rebello, M. Smart, C. Wang, K. Czarnecki, and S. Waslander, “Canadian Adverse Driving Conditions Dataset,” *The International Journal of Robotics Research*, 2020.
- [32] P. Sun *et al.*, “Scalability in Perception for Autonomous Driving: Waymo Open Dataset,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020.
- [33] D. Barnes, M. Gadd, P. Murcutt, P. Newman, and I. Posner, “The oxford radar robotcar dataset: A radar extension to the oxford robotcar dataset,” in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 6433–6438.
- [34] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, “nuScenes: A Multimodal Dataset for Autonomous Driving,” in *(CVPR)*, 2020.
- [35] M. Meyer and G. Kuschik, “Automotive Radar Dataset for Deep Learning Based 3D Object Detection,” in *European Radar Conference (EuRAD)*, 2019.
- [36] Y. Schulz, A. K. Mattar, T. Hehn, and J. Kooij, “Hearing What You Cannot See: Acoustic Vehicle Detection Around Corners,” *Robotics and Automation Letters*, 2021.
- [37] J. Binas, D. Niel, S.-C. Liu, and T. Delbruck, “DDD17: End-To-End DAVIS Driving Dataset,” in *ICML’17 Workshop on Machine Learning for Autonomous Vehicles (MLAV 2017)*, 2017.
- [38] A. Z. Zhu, D. Thakur, T. Özarslan, B. Pfrommer, V. Kumar, and K. Daniilidis, “The Multivehicle Stereo Event Camera Dataset: An Event Camera Dataset for 3D Perception,” *Robotics and Automation Letters*, vol. 3, no. 3, 2018.
- [39] Z. Zhang and D. Scaramuzza, “A tutorial on quantitative trajectory evaluation for visual (-inertial) odometry,” in *(IROS)*. IEEE, 2018, pp. 7244–7251.
- [40] R. B. Rusu and S. Cousins, “3D is here: Point Cloud Library (PCL),” in *IEEE International Conference on Robotics and Automation (ICRA)*, Shanghai, China, May 9-13 2011.
- [41] W. Lewandowski and C. Thomas, “GPS time transfer,” *Proceedings of the IEEE*, vol. 79, no. 7, pp. 991–1000, 1991.
- [42] A. Savitzky and M. J. Golay, “Smoothing and differentiation of data by simplified least squares procedures.” *Analytical chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964.
- [43] “camera_calibration - ROS Wiki,” 2021. [Online]. Available: http://wiki.ros.org/camera_calibration#Camera_Calibrator