

CMU-GPR Dataset: Ground Penetrating Radar Dataset for Robot Localization and Mapping

Alexander Baikovitz¹, Paloma Sodhi¹, Michael Dille², Michael Kaess¹

Abstract—There has been exciting recent progress in using radar as a sensor for robot navigation due to its increased robustness to varying environmental conditions. However, within these different radar perception systems, ground penetrating radar (GPR) remains under-explored. By measuring structures beneath the ground, GPR can provide stable features that are less variant to ambient weather, scene, and lighting changes, making it a compelling choice for long-term spatio-temporal mapping. In this work, we present the CMU-GPR dataset—an open-source ground penetrating radar dataset for research in subsurface-aided perception for robot navigation. In total, the dataset contains 15 distinct trajectory sequences in 3 GPS-denied, indoor environments. Measurements from a GPR, wheel encoder, RGB camera, and inertial measurement unit were collected with ground truth positions from a robotic total station. In addition to the dataset, we also provide utility code to convert raw GPR data into processed images. This paper describes our recording platform, the data format, utility scripts, and proposed methods for using this data.

I. INTRODUCTION

We present the CMU-GPR dataset—the first open-source ground penetrating radar (GPR) dataset available to researchers interested in subsurface-aided perception for robot navigation, to the best of our knowledge. Radar-based perception has been shown to perform more robustly than conventional spatial or visual sensors in inclement weather [1–3]. While recent work has focused on using millimeter-wave radar systems to construct surface-level models, there has been growing interest in using subsurface information from GPR for localization [4, 5].

GPR presents a modality to recognize a location in situations where the visual environment may change by perceiving typically constant subsurface features. For instance, a robot operating in a mining environment may encounter substantial changes in the surface environment due to mining operations, yet underground features remain more consistent. Similarly, GPR-based localization can be effective in sparsely featured environments, such as monotonous tunnels and open roads, where subsurface geologic diversity can enable lane tracking with respect to a prior map.

In this contribution, we produce a dataset and utility functions to empower researchers to explore how GPR can be used as a tool for robot navigation applications. The dataset collected contains subsurface measurements from a low-cost, off-the-shelf, single-channel GPR system along with a wheel encoder, IMU, RGB camera, and robotic total station.

The CMU-GPR dataset and related utility functions are available at <https://github.com/rpl-cmu/CMU-GPR-Dataset>. This work was supported by a NASA Space Technology Graduate Research Opportunity.

¹ The Robotics Institute, Carnegie Mellon University, ² Intelligent Robotics Group, KBR / NASA Ames Research Center, Correspondence to abbaikovitz@cmu.edu

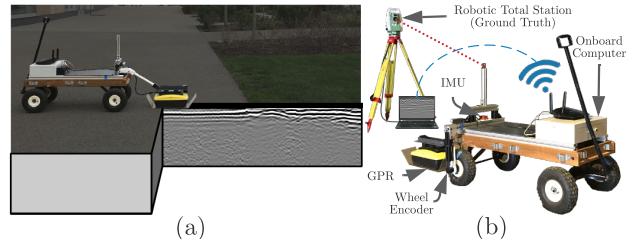


Fig. 1: SuperVision platform for collecting subsurface data using GPR sensor during motion. (a) A cross section of the subsurface is shown containing different flat boundaries and a pipe. (b) Schematic of the SuperVision platform used for data collection.

II. THE SUPERVISION PLATFORM

SuperVision shown in Figure 1(b) is a custom, manually-pulled experimental rig used for acquiring subsurface data from GPR. SuperVision uses a quad-core Intel NUC for compute and wirelessly transmits onboard data from an XSSENS MTI-30 9-axis Inertial Measurement Unit, YUMO quadrature encoder with 1024 PPR, Intel RealSense D435 camera, and a Sensors and Software Noggin 500 GPR. Ground truth data was acquired by a Leica TS15 total station. The base station logs measurements from both the onboard computer and total station to ensure time synchronization.

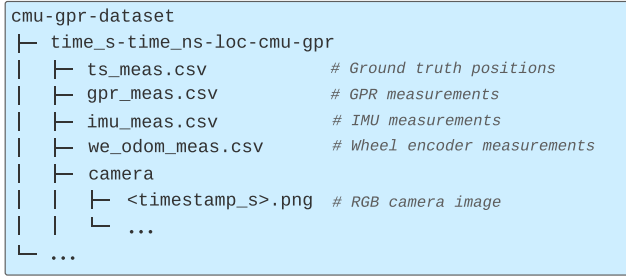
III. DATA COLLECTION

The CMU-GPR dataset contains short trajectories from three distinct, GPS-denied environments: a basement (*nsh_b*), a factory floor (*nsh_h*), and a parking garage (*gates_g*). The dataset consists of 15 traversals where the manually-pulled experimental rig revisits previous locations, whether through forward-backward motion or by closing loops. Several distinct trajectories contain similar features, which are relevant to research on re-localization using subsurface information and further described on the project website. Additionally, we provide trajectories from outdoor environments that do not contain ground truth information, which present additional data to train models.

Figure 2(a) shows the directory structure for a single trajectory sequence. Each zip file contains data from a sequence, where each *.csv* represents a different measurement type. The maximum size of a zipped dataset file is 4 GB. The formats for each measurement type are summarized in Figure 2(b). In addition to the data, the project website contains relevant IMU noise and bias parameters as well as the factory extrinsic calibration of the system.

IV. DEVELOPMENT TOOLS

Along with relevant datasets for localizing GPR, we provide utility code, written in Python, which processes



(a)

Measurement Type	Measurement CSV Value
GPR Measurement:	$[t_{\text{stamp}}, \underbrace{\text{amp}_1, \dots, \text{amp}_{201}}_{\text{Amplitude of 1D Trace}}]$
IMU Measurement:	$[t_{\text{stamp}}, \underbrace{a_x, a_y, a_z}_{\text{Accelerometer [m/s}^2\text{]}}, \underbrace{g_x, g_y, g_z}_{\text{Gyroscope [rad/s]}}, \underbrace{w, x, y, z}_{\text{Magnetometer}}]$
Wheel Encoder Measurement:	$[t_{\text{stamp}}, \underbrace{\text{dist}_x}_{\text{Signed distance traversed [m]}}, \dots]$
Total Station Measurement:	$[t_{\text{stamp}}, \underbrace{p_x, p_y, p_z}_{\text{Ground truth position [m]}}, \dots]$

(b)

Fig. 2: (a) Directory layout for the CMU-GPR dataset. (b) Data format by sensor measurement type.

the raw GPR data to construct images. A modular routine to process 1D measurements and construct an image with uniform spacing is described in Section IV-A. Additionally, we provide a script to generate submaps, which can be used for training or evaluating GPR sensor models.

A. Signal Processing

Utility code to process raw GPR signals and images is provided in `signal_processing_utils.py` and used in `metric_gpr_image.py`. The implementation accepts raw, unevenly spaced GPR images and produces processed brightness scans. The base pipeline performs rubber band interpolation, mean background subtraction, dewow filtering, triangular bandpass filtering, zero time correction, SEC gain, wavelet denoising, and gaussian filtering.

B. Image Construction

Beyond processing the raw data, the utility code also simplifies image acquisition. The `MetricGprImage` object stores the GPR dataset and allows the client to access images based on the time of acquisition. The `ImageConstructor` object is even more abstract, allowing the client to create a traditional radargram of the entire sequence or automatically generate all valid submap images.

V. APPROACHES

The dataset proposed in this paper was used for robot localization in an unknown, GPS-denied environments using a GPR sensor. In that work, Baikovitz *et al.* [6] utilized learned sensor models to incorporate GPR submaps into a factor graph-based estimation framework. These learned models provided relative motion predictions during loop closures thereby correcting for accumulated drift in the

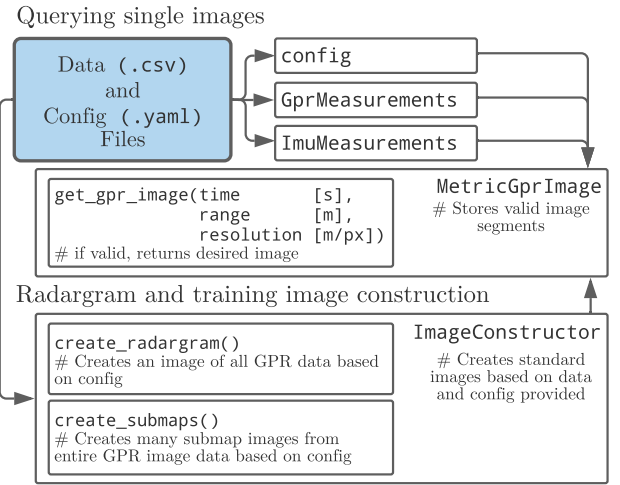


Fig. 3: API for constructing processed images from CMU-GPR dataset.

position estimates. The goal of this work was to demonstrate how a low-cost, off-the-shelf, single-channel GPR system can be used effectively for robot localization.

We see many avenues for extending the use of this dataset and GPR in robotics. One limitation with a localization-only approach is that one needs to revisit prior locations in order to correct for drift. This can be avoided if we maintain an online map of the subsurface structures such as pipes, mines, and natural caves, and localize with respect to it. This would involve solving a simultaneous localization and mapping (SLAM) problem. Finding the correct representation for GPR data is likely the largest challenge for robust, GPR-based SLAM. We observe that a 2D image representation performs well compared to raw 1D traces, which do not provide unique enough features, and 3D point clouds, which are sensitive to noise and variable subsurface composition. Some prior work addresses automatic detection of hyperbolic features in 2D GPR images; however, these methods are often heuristic-based and not investigated for use in a SLAM graph [7].

VI. DISCUSSION

Our motivation for providing this contribution is to encourage others in the field to take similar steps in making GPR-based perception datasets available to researchers. We believe that providing this data to the research community will spur further development of robust GPR-based localization systems for real world deployment.

REFERENCES

- [1] 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 *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, Paris, 2020.
- [2] M. Sheeny, E. De Pellegrin, S. Mukherjee, A. Ahrabian, S. Wang, and A. Wallace, "Radiate: A radar dataset for automotive perception," *arXiv*, 2020.
- [3] Y. S. Park, J. Jeong, Y. Shin, and A. Kim, "Radar dataset for robust localization and mapping in urban environment," in *ICRA Workshop on Dataset Generation and Benchmarking of SLAM Algorithms for Robotics and VR/AR*, 2019.
- [4] M. Cornick, J. Koehling, B. Stanley, and B. Zhang, "Localizing Ground Penetrating RADAR: A Step Toward Robust Autonomous Ground Vehicle Localization," *Journal of Field Robotics*, vol. 33, no. 1, pp. 82–102, jan 2016.
- [5] 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.
- [6] A. Baikovitz, P. Sodhi, M. Dille, and M. Kaess, "Ground encoding: Learned factor graph-based models for localizing ground penetrating radar," in *arXiv*, 2021.
- [7] Q. Dou, L. Wei, D. R. Magee, and A. G. Cohn, in *Real-Time Hyperbola Recognition and Fitting in GPR Data*, 2017.