Test Your SLAM! The SubT-Tunnel dataset and metric for mapping

John G. Rogers III[1], Jason M. Gregory[1], Jonathan Fink[1], Ethan Stump[1]

Abstract— This paper presents an approach and introduces new open-source tools that can be used to evaluate robotic mapping algorithms. Also described is an extensive subterranean mine rescue dataset based upon the DARPA Subterranean (SubT) challenge including professionally surveyed ground truth. Finally, some commonly available approaches are evaluated using this metric.

I. INTRODUCTION

Situational awareness in a dynamic and evolving disaster recovery scenario is a critical component in allocating resources to minimize survivor casualties. In the subterranean regime, recovery can be hampered by limited access due to collapse or fire, as well as environmental hazards such as noxious fumes and the risk of further collapse. Robots can be employed to locate survivors and build maps which enable incident commanders to make decisions balancing exposure to risk with reducing human casualties. To enable roboticists to speak to these challenges, we present the **SubT-Tunnel** dataset and analysis tools intended for benchmarking simultaneous localization and mapping (SLAM) algorithms in underground tunnel environments. This dataset represents a snapshot of the Tunnel Circuit and SubT Integration eXercise (STIX) events of the DARPA Subterranean Challenge[2], held in the United States at the National Institute for Occupational Safety and Health (NIOSH) and Edgar Mines in Pennsylvania and Colorado, respectively. Mines present a starkly different set of environment characteristics than those found in typical SLAM datasets: poor to no lighting, varied levels of roughness and irregularity in structure, sometimes significant changes in topography, wet, dirty, and no access to GPS. By providing a dataset which exposes these characteristics, we aim to broaden the classes of environments that SLAM algorithms should be judged against.

In designing this dataset, we have followed the structure of other datasets, such as the KITTI Vision Benchmarking Suite [3], that contain a redundant set of sensory inputs in order to admit multiple algorithmic techniques. In addition to the standard stereo camera imagery and multiple LiDAR sources, we also include thermal camera data as an additional modality that provides special utility in low-light, underground conditions. This style of benchmark dataset enables ablation studies, where algorithmic components and sensor inputs are switched off or degraded to objectively determine marginal effects on performance.

[1]CCDC Army Research Laboratory (ARL), Adelphi, MD 20783, USA. {john.g.rogers59, jason.m.gregory1, jonathan.r.fink3, ethan.a.stump2}.civ@mail.mil
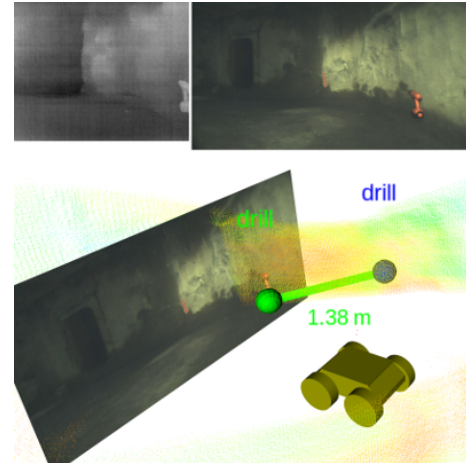
[2]https://subtchallenge.com

Fig. 1. A sample landmark artifact report visualization resulting in a scored point. The robot, shown in yellow, is positioned within the NIOSH Experimental mine at the DARPA Subterranean Challenge. The drill (orange) is visible in the robot's stereo camera (upper right) as well as thermal IR camera (upper left). The ground truth position of the drill in a global reference frame is shown by the blue sphere, while the detection report is shown with the green sphere. As the detection is within the scoring threshold, the reported location is shown in green which indicates a point is scored. Additionally, RMSE is computed on the artifact reports to allow for fine-grained comparison and evaluation of mapping accuracy.

Obtaining ground truth information, namely the actual sensor trajectory, for SLAM in extensive environments is challenging, and many datasets only provide complete ground truth for a small environment, partial ground truth through a small portion of a larger traversal, or simply rely on users doing qualitative analysis on the resulting map. We take the approach of providing ground truth modeled after the scoring of the DARPA Subterranean Challenge: during each collection, we have manually labeled the detection and localization of several competition "artifacts," and we can evaluate the SLAM accuracy by composing these locations with the estimated position of the platform and comparing against the surveyed artifact locations. The **SubT-Tunnel** dataset thus represents a simplified version of the competition.

In terms of mapping analysis, there are two general methodologies for measuring mapping accuracy, relative and absolute [8], and we fall squarely on the side of absolute error for this application. This puts much more stringent requirements on the mapping system but is motivated by the observation that many external interactions with the tunnel, such as drilling from above to ventilate or provide emergency supplies to survivors, can only be successful with accurate absolute estimates. In addition, by focusing on a small set of target locations, we are suggesting a shift to a *task-based*

benchmarking mindset for SLAM. This not only lowers the cost of obtaining good ground truth, but also allows the analysis to apply to a wider range of possible mapping sources, including multiple agents working with independent maps but a shared coordinate system.

The contributions of this paper are as follows:

1) We introduce and describe a new, open dataset taken from the DARPA Subterranean Challenge Tunnel Circuit competition and STIX events.
2) We propose a new absolute-accuracy analysis metric for map evaluation modeled after the competition scoring.
3) We provide a set of open-source support tools to enable researchers to easily evaluate their own mapping approaches against this metric, including an approach for aligning their map coordinate frames with the global frame without using GPS.
4) As a baseline, we compare multiple SLAM algorithms from the literature using this metric.

In doing so, we seek to provide a useful resource for researchers looking to get started in understanding the challenges of subterranean operations.

II. Data Collection

The **SubT-Tunnel** dataset[3] consists of sensor information from two mine sites. One recording was taken at the Edgar research mine, a former precious metals mine, in Idaho Springs, Colorado during the DARPA STIX, courtesy of the Colorado School of Mines. The other recordings are from the mine used for the DARPA Subterranean Challenge Tunnel circuit in Bruceton, Pennsylvania, courtesy of NIOSH. The mines used in this dataset were shut-down over 100 years ago and have since been re-opened for educational, research, and training purposes by the aforementioned parties. These locations were chosen by DARPA to provide representative examples of mine rescue, while maintaining safety for participants and personnel. Each mine presents some specific mapping challenges:

- *STIX*: The Edgar mine has a few improvements to illumination, but is dark along most of its tunnels. The ground is mostly flat, with some limited portions with steep inclines. The mine also features a prominent railroad track over significant portions that makes mobility challenging for ground robot platforms. The mine is dry enough that the robot produces significant atmospheric dust at times that may scatter LiDAR and obscure visible light cameras.
- *Tunnel Circuit*: The Bruceton mine is divided into two separate courses, Safety Research (SR) and Experimental (EX). Illumination was provided by sparse overhead lighting but does feature some dark areas. The ground is mostly flat but wet and muddy, with flooded areas that reflect LiDAR in addition to visible light.

---

[3]Instructions for downloading the SubT-Tunnel dataset and the supporting code repository described in this paper can be found at https://bitbucket.org/jgrogers/subt_reference_datasets
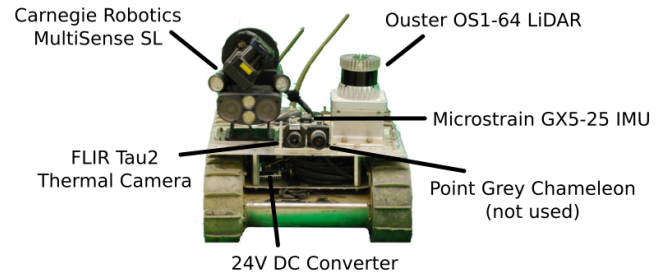

Fig. 2. Robot used to collect data at the "Edgar" mine during the STIX event.

Each of the tunnel circuit challenge courses, SR and EX, are given in two configurations: A, and B (only configuration B is represented in this dataset). For each of these configurations, 20 artifacts are placed within the courses. As teams achieve points by accurately reporting the locations of these artifacts, accurate ground truth positions for the artifacts are necessary for comparison. To achieve the required level of accurate position information for the artifacts, the ground truth positions were determined by professional surveyors experienced in mine surveying.

A. Sensors and Robots

For the STIX dataset, the iRobot PackBot was chosen due to its tracked configuration to improve maneuverability over the railroad tracks of the Edgar Mine. The robot and its sensor payload are shown in Figure 2. This robot is equipped with an Ouster OS1-64 to collect the primary high-density LiDAR point cloud. A second LiDAR sensor is also included in the Carnegie Robotics MultiSense SL for redundancy and comparative purposes. The Multisense SL sensor also provides stereo vision and high-intensity illuminators, which are necessary to make observations of artifacts in the darker portions of the environment. The robot is also equipped with a FLIR Tau2 thermal infrared camera since several of the artifacts are heated (mock survivors and cell phones) and are quite apparent on thermal vision. The dataset provides platform odometry that is generated using wheel odometry with orientation replaced from a Microstrain 3DM-GX5-25 IMU. The Received Signal Strength Indicator (RSSI) to all available wireless 2.4GHz hotspots is also recorded. This is primarily useful in locating the cell phone artifacts; secondarily, it is planned for use in subterranean radio propagation analysis.

For the Tunnel Circuit datasets, a Clearpath Husky was chosen for its superior ground clearance and rugged offroad tires. The robot platform, complete with deployed sensor payload, can be seen in Figure 3. The robot is equipped with the same sensor payload as used in the STIX datasets with the following modifications. The Ouster OS1-64 LiDAR has been moved to a location on the robot where its view will not be occluded by other components. The other primary change is that the FLIR Tau2 has been replaced with a FLIR Boson camera as the frame-grabber for the Tau was not reliable when collecting data at the STIX event, and the Boson camera has a direct USB interface board.
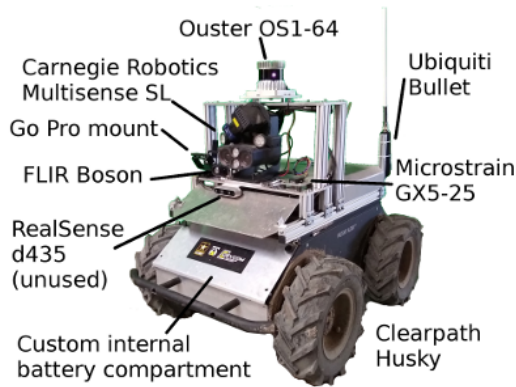
Fig. 3. Robot used to collect data at the SR and EX courses from the tunnel circuit competition mine.

All data collection was done under direct teleoperation, with the operator following behind the robot. For the STIX data, an XBox 360 wireless controller was used that has limited range; therefore, personnel are occasionally visible in the LiDAR view behind the robot. For the tunnel circuit data, teleoperation control was managed over a WiFi link that enabled the operators to stay further back and avoid being seen by the robot. Be aware that the operators are detected by the LiDAR at times, but if the full scan is used, these errors are relatively minor.

B. Dataset Collection

The dataset is collected in Robot Operating System (ROS[4]) bag format using the tool *rosbag record*. Some sensor output is recorded both in its raw datagram packet form, and in some convenient conversions to more immediately usable forms such as point clouds. Camera imagery is recorded using *image_transport* compressed type to greatly reduce file size while maintaining full frame rate and similar quality. Example ROS launch files demonstrating how to play back this dataset can be found in the *tunnel_ckt_launch* directory of the included repository referred to in Section II.

Many more trajectories were recorded than appear in this description. A faulty encoder on the Husky rendered odometry unusable on Configuration A runs at the tunnel circuit. These runs were therefore not evaluated here as most of the techniques we evaluate are configured to rely upon odometry. One run with poor odometry, *ex_B_route2.bag*, was included in the evaluation to see how well these techniques could perform with poor odometry. The runs in Configuration A with poor odometry are expected to be released in the future or can be provided by request, as they could still be useful to evaluate visual SLAM.

III. METRICS FOR COMPARISON

In prior SLAM metric comparison work such as [1], an incremental relative error was chosen due to the assumption that an equivalent error made early on in a map should result in the same "score" as an error made much later in the map.

This relative error metric is useful for many applications; however, in the subterranean regime, minimizing absolute error is needed in many cases. For example, in a worst-case mine collapse rescue, an access tunnel might need to be drilled from the surface. To hit the desired passageway requires accurate absolute position, within the characteristic width of the tunnel (in our case, this is around 2.5 meters). In the case of reporting the location of survivors to first responders, a report made within 5 meters of the survivor would likely be close enough for the first responder to find that survivor.

Since the ground truth survey gives us accurate absolute positions of the artifacts that are distributed throughout the challenge courses, they can be used as landmarks to compare against to establish accuracy for a mapping procedure. The first metric is the number of artifacts observed within 5 meters of their ground truth positions, which corresponds to the score a team would have received had they used their SLAM algorithm and their robot had driven the same path through the course. As this is a relatively coarse measure, the Root Mean Square Error (RMSE) on the distance to the ground truth artifact landmarks is also used to give a fine-grained comparison. Finally, the minimum error and maximum error to the artifact landmark position are also given to establish the stability of the mapping procedure over the length of a run.

To support testing of SLAM systems independent of object recognition, artifact locations are coded into run files with the extension *àrtifacts*. These artifact locations are generated through the *subt_scoring_node* with the *coding_mode* parameter set to *true*. In coding mode, the scoring node generates the artifact file based upon user input. The location of AprilTags corresponding to the ground truth frame origin are generated through the use of the AprilTag library, which is provided as an entry in the included rosinstall file. As the AprilTags for the fiducial landmarks are passed, the best measurement as reported by the tag detection is kept and recorded into the àrtifacts file. The entry in the artifacts file consists of the fiducial tag, the ROS time when the detection was made, the image coordinates of the detection (center), the camera frame ID, and 3D coordinates of the tag (using stereo for depth) and the robot base frame ID and 3D coordinates in the base frame. When the user sees that the robot is observing an artifact that should be coded into the file, they first pause the bag playback, and then click the button on the right of the coding window corresponding to the artifact type. The user then clicks on the center of the artifact in the image shown in the window. The same type of entry as described for fiducial landmarks is made in the artifacts file, except using the artifact type string.

Since the artifact locations are recorded in a local coordinate system, the scoring node simply looks for a transform between the darpa frame (established by the ground truth locations of the fiducial landmarks) and the local frame. As the darpa to map frame is automatically provided by the scoring node, the user needs only to provide the map to "chinook/base" portion of the transform in TF. The user
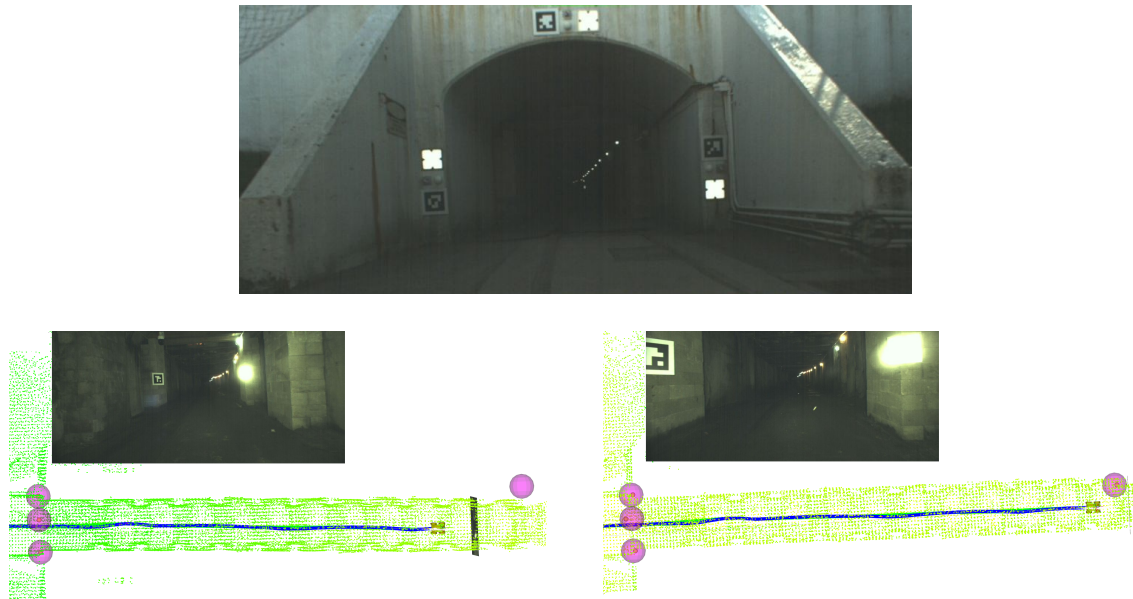
[4]www.ros.org

957

Fig. 4. Top: View from the robot of the entrance portal and the frame alignment fiducials. These AprilTags, along with IR reflectors, prisms, and spheres, have been surveyed in the "darpa" reference frame. This information is meant to be used by competitors to find the transform to this frame as their robots enter the challenge. Bottom left: A top-down view of the robot after it has entered the tunnel, but before the distal fiducial tag is used to improve the alignment of the "darpa" to map frame. Bottom right: The same view taken moments later after the distal fiducial is used to correct the absolute frame transform. In both of these images, the large pink spheres represent the fiducial location in the global frame, and the smaller spheres represent the measurements in the map frame.

can include the "chinook/odom" to chinook/base and opt to provide only the correction in map to chinook/odom. Using this transform, the scoring node computes the darpa frame location of the landmark that would have resulted from making this observation while using the mapping system under evaluation. The relationships between these frames are summarized in Figure 5. The scoring node processes the artifacts file and compares artifact positions to the ground truth when the ROS time of the report creation is reached. If the location is within 5 meters of the ground truth position, a point is accumulated. The RMSE and min/max errors are also updated to reflect this new measurement.
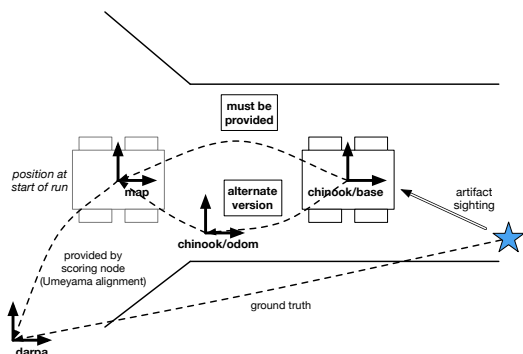

Fig. 5. TF frame information; analysis is performed on the transform from *chinook/base* to *map* that a SLAM algorithm generates.

Alignment of the mapping system under test to the global darpa frame is handled automatically by the scoring node. The procedure is illustrated in Figure 4. The AprilTags were observed in the coding node, and the most confident match

is recorded along with the position of the tag in the robot's base frame. The stereo depth is currently used to find the distance to the tags. Upon observing a minimum of three tags, the transform between the darpa and map frames is estimated using the Umeyama algorithm [10] provided in the Eigen library.

This dataset could be used to additionally test an object recognition and localization mechanism; this would require an extension to the scoring node to substitute artifact reports processed from the coded artifacts file with a report callback (which would have to be added to the scoring node). Care should be taken to limit the number of reports made on a single artifact; the artifacts were coded only on one appearance in a given "loop" (i.e. the robot had to travel a significant distance before revisiting the artifact for it to be subsequently recoded).

IV. BENCHMARKS

We have selected a set of modern SLAM approaches that were available in the open source in addition to our own technique. These approaches were selected to provide some examples of using this dataset with different sensory modalities, including LiDAR and stereo vision. In addition, the raw proprioceptive odometric trajectory estimate is evaluated by itself to establish how necessary mapping is to achieving a high score on this challenge.

OmniMapper

The first SLAM algorithm evaluated with this dataset is an approach from our previous work that was developed based upon the open-source *OmniMapper* [9]. This mapping system is a modular framework for integrating sensory mea-

surements from potentially many modalities. It is configured for this analysis to utilize platform wheel odometry, gyro data for orientation only, and LiDAR data from the Ouster OS1-64. OmniMapper builds a *pose graph* along the robot's trajectory, with connections between adjacent poses coming from Iterative Closest Point (ICP) [7] when a match is made with low residual error, and wheel odometry is substituted when no match is found. Wheel odometry is used as an initial guess to bootstrap the ICP iterations, which greatly accelerates convergence. When the robot revisits a location previously seen along its trajectory, the ICP procedure is used to find *loop closures*, where additional constraints are added to the graph. This graph of relative frame transformations is continuously optimized via iSAM2 [5] in the GTSAM package [2]. The dataset evaluations are performed in real time; however, only partially degraded performance is still achieved at 400% speed.

Cartographer

Cartographer [4] is an open-source mapping suite which can incorporate 2D and 3D LiDAR, landmark measurements, odometry, and IMU measurements. In place of ICP, *Cartographer* uses a fast correlation-based scan matcher for both frame-to-frame matching as well as loop closure. These measurements are incorporated into a graphical representation that is solved by the *Ceres* solver.

For our analysis, we have included the launch files and configurations used for *Cartographer*. These parameters are not necessarily ideal and platform calibrations might not be up to the level that is needed for this software. In some of the runs made with *Cartographer*, the robot appears to descend and gets the wrong altitude early in the run, which limits the scoring accuracy. As this is likely due to a platform calibration issue or configuration parameter, we have included an alternative analysis where the artifact locations and robot detections are projected into the X-Y plane, effectively eliminating this potential source of error. These results can be found in the *Cartographer_2D* entries in Table I.

ORB_SLAM2

ORB_SLAM2 [6] is a robot mapping system that consumes camera data instead of LiDAR, either monocular, stereo, or depth camera (RGB-D). This approach leverages *visual odometry* to build local keyframe models. Bundle adjustment is used to triangulate the location of landmarks across keyframes. When locations are revisited, a bag-of-words description is used to identify loop closures. This mechanism is also used to provide a reliable relocalization mechanism in the event of tracking failure. This approach has been evaluated against many existing (above ground) mapping datasets and has been demonstrated to achieve state-of-the-art accuracy.

Experimental Evaluation. Each of the techniques described previously was evaluated against the dataset where possible. Qualitative results in the form of point clouds overlaid on ground-truth-surveyed point cloud data can be seen in Figure 6. In this figure, both of the tunnel circuit courses can be seen. The *safety research* course is the top section

of the mine, and the *experimental* course is in the bottom section of the mine.

Results for each tested algorithm can be seen in Table I. In this table, a dash indicates that the algorithm was unable to be run on the dataset. The *Cartographer* algorithm (in both 3D and 2D configurations) was not run on the data taken in the STIX event at the Edgar mine. This is due to a mistake on the data collection, where the necessary odometry message was omitted. As this would deprive the *Cartographer* algorithm of one of its key inputs, it was decided to omit the results for this dataset. *Cartographer* can indeed function without the odometry input; however, its accuracy with only IMU and laser scan data was insufficient to come close to scoring any points. We believe that this may be due to imprecise calibration of IMU; this algorithm might be able to achieve significantly better results with ideal calibration. Overall, the *OmniMapper* performs slightly better than *Cartographer*; however, this may be attributed to the fact that *OmniMapper* has been tuned specifically for this operation and we may not have configured *Cartographer* ideally.

ORB_SLAM2 was not configured to run on the STIX course since the right camera image was omitted. The depth image is available, so results are anticipated in the future using the RGB-D configuration. The main issue with *ORB_SLAM2* on this dataset is that, due to the darker conditions in the mines, feature tracking fails when the robot rotates at even a moderate rate. The relocalization procedure in *ORB_SLAM2* was often unable to recover, as the robot does not return to revisit these locations until much later in the run. We modified the recovery procedure to continue with the last pose when tracking was lost, which is designated by *ORB_SLAM2+*. As the vehicle was often rotating when tracking was lost, this procedure enabled the mapping run to continue but typically introduced significant tracking error. In some cases, additional progress was made; however, no additional artifacts were identified with sufficient precision to count for additional score. Each of the *ORB_SLAM2* runs was terminated when the robot moved past the lit portion of the mine, where it became totally non-functional, despite the illumination provided by the onboard illuminators.

The illumination in this dataset is clearly less than what is typically given to *ORB_SLAM2*; however, by observation the illuminators clearly provide enough light for an observer to determine the platform's motion. It remains to be demonstrated if visual SLAM-based techniques can operate at the precision needed to score well on the Subterranean Challenge; the authors plan to test other approaches in the future.

V. Conclusion and Future Work

We have presented a dataset, a metric, and analysis tools for evaluating mapping algorithms applied to underground tunnel environments. Thanks to the efforts of the DARPA Subterranean Challenge support team, surveyed ground truth landmark artifact positions are provided that extend deep below the earth in three mines which has enabled us to challenge researchers to test their own mapping systems. We
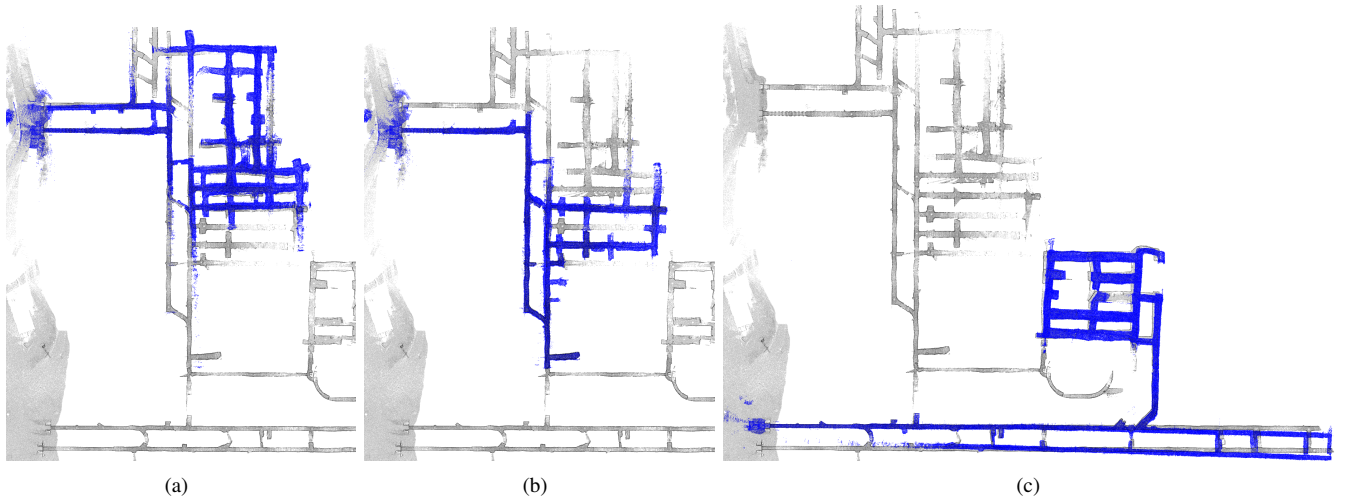
Fig. 6. Ground truth point clouds of the tunnel environment by professional surveying are shown in black. OmniMapper results are overlaid in blue for qualitative comparison.

Run		Algorithm	Score	RMSE [m]	Min err [m]	Max err [m]
File	Length					
sr_B_route1.bag	909	OmniMapper	8/8	1.12	0.31	2.22
		Cartographer	7/8	3.3	1.2	7.2
		Cartographer_2D	8/8	1.5	0.71	2.1
		ORB_SLAM2+	1/2*	13.7*	1.2	19.2
		ORB_SLAM2	1/1*	1.2	1.2	1.2
		Odometry	1/8	13.25	4.44	27.7
sr_B_route2.bag	792	OmniMapper	9/9	1.5	0.46	2.7
		Cartographer	9/9	2.3	1.7	3.0
		ORB_SLAM2+	1/5*	28.3	2.8	51.6
		ORB_SLAM2	1/2*	4.8	2.8	6.2
		Odometry	0/7	9.6	5.5	15.3
ex_B_route1.bag	1930	OmniMapper	20/20	2.4	0.95	4.1
		Cartographer	1/20	11.9	4.4	18.2
		Cartographer_2D	18/20	2.6	0.23	6.7
		ORB_SLAM2+	3/6*	15.1	1.7	32.2
		ORB_SLAM2	3/3*	2.62	1.7	3.5
		Odometry	0/20	54.9	5.29	168.1
ex_B_route2.bag	1187	OmniMapper	9/13	9.1	0.26	22.1
		Cartographer	4/13	19.4	2.23	46.2
		ORB_SLAM2+	0/0*	0	0	0
		Odometry	4/13	23.8	2.2	44.3
stix_mainloop.bag	871	OmniMapper	7/19	14.0	0.27	37.2
		Cartographer	-	-	-	-
		ORB_SLAM2	-	-	-	-
		Odometry	4/19	19.3	0.53	47.8

TABLE I

MAPPING SCORE, RMSE, MINIMUM, AND MAXIMUM ERROR FOR EACH ALGORITHM ON THE DATASET

have performed an initial evaluation on our own mapping system, and two options available in the open source, as well as an evaluation of how well the robot's proprioceptive sensing alone could be used in place of mapping.

Our proposed metric focuses on the absolute accuracy of a small set of locations rather than the accumulated trajectory error that is often reported. We argue that this is valid because it represents a *task-based benchmarking* mindset: we are interested in assessing how well mapping enables the mission functions of the robotic system. More importantly, this provides a pathway for analyzing systems that look drastically different than the "single moving sensor package" that describes essentially every SLAM dataset.

One advantage of a modern graph-based SLAM approach is that localization along the trajectory can be corrected through loop closure. Our analysis methodology described here scores a point each time an observation of an artifact is made. In the future, one improvement to this technique would be to incorporate this trajectory correction by using the final posterior estimate instead of the intermediate one. The methodology described here is one that can be easily applied in many situations; future work will focus on enabling these type of posterior artifact reports in a universally applicable way.

Though this initial analysis used hand-coded artifact detections, this dataset has the potential for evaluating object recognition and localization alongside the current map accuracy evaluation if users wanted to run their own perception algorithms. A mechanism could be implemented in the scoring node that would accept global frame object reports and score them directly against the ground truth; replacing the coded artifact reports. We invite the community to submit pull requests to the repository that implement support for this mode of evaluation.

We intend to create similar datasets at each of the future DARPA Subterranean Challenge circuit events as well as the final combined event. The future circuit events are planned to extend beyond mines; the next circuit event is planned to involve the "Urban Underground", the infrastructure beneath the city streets. The final circuit event is planned to take place in natural caves. As each of these planned future circuit events has the potential to involve terrain that is inaccessible to a wheeled platform, we plan to add aerial assets to supplement future collections.

VI. ACKNOWLEDGEMENT

References

[1] W. Burgard, C. Stachniss, G. Grisetti, B. Steder, R. Kümmerle, C. Dornhege, M. Ruhnke, A. Kleiner, and J. D. Tardös. A comparison of SLAM algorithms based on a graph of relations. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2089–2095. IEEE, 2009.

[2] F. Dellaert. Factor graphs and GTSAM: A hands-on introduction. Technical report, Georgia Institute of Technology, 2012.

[3] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun. Vision meets Robotics: The KITTI Dataset. *Int. Journal of Robotics Research*, 2013.

[4] W. Hess, D. Kohler, H. Rapp, and D. Andor. Real-Time Loop Closure in 2D LIDAR SLAM. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1271–1278, 2016.

[5] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. Leonard, and F. Dellaert. iSAM2: Incremental smoothing and mapping with fluid relinearization and incremental variable reordering. In *2011 IEEE International Conference on Robotics and Automation*, pages 3281–3288. IEEE, 2011.

[6] R. Mur-Artal and J. D. Tardós. ORB-SLAM2: An open-source SLAM system for monocular, stereo, and RGB-D cameras. *IEEE Transactions on Robotics*, 33(5):1255–1262, 2017.

[7] A. Segal, D. Haehnel, and S. Thrun. Generalized-ICP. In *Robotics: science and systems*, volume 2, page 435. Seattle, WA, 2009.

[8] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers. A benchmark for the evaluation of RGB-D SLAM systems. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 573–580. IEEE, 2012.

[9] A. J. Trevor, J. G. Rogers, and H. I. Christensen. Omnimapper: A modular multimodal mapping framework. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1983–1990. IEEE, 2014.

[10] S. Umeyama. Least-squares estimation of transformation parameters between two point patterns. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (4):376–380, 1991.