

UPSLAM: Union of Panoramas SLAM

Anthony Cowley, Ian D. Miller and Camillo Jose Taylor

Abstract— We present an empirical investigation of a new mapping system based on a graph of panoramic depth images. Panoramic images efficiently capture range measurements taken by a spinning lidar sensor, recording fine detail on the order of a few centimeters within maps of expansive scope on the order of tens of millions of cubic meters. The flexibility of the system is demonstrated by running the same mapping software against data collected by hand-carrying a paired lidar and IMU around a laboratory space at walking pace, moving them outdoors through a campus environment at running pace, driving the sensors on a small wheeled vehicle on- and off-road, flying the sensors through a forest, carrying them on the back of a legged robot navigating an underground coal mine, and mounting them on the roof of a car driven on public roads. The full 3D maps are built online with a median update time of less than ten milliseconds on an embedded NVIDIA Jetson AGX Xavier system.

I. INTRODUCTION

Simultaneous localization and mapping (SLAM) is a core capability underpinning general field robotics. It is a keystone for software needing to relate instantaneous readings from a robot's sensors to historical data – anything from what the robot perceived a fraction of a second in the past, to a pre-loaded map of a location – while allowing the robot to extend its corpus of knowledge about how its sensors perceive its surroundings. Relating current sensor data to a map permits localization, while computing transformations that explain how sensor data changes over time enables the extension of that map with new data. The degree to which existing techniques meet the challenge of SLAM depends upon factors such as [1]:

- **Motion** Will the sensor be moved quickly or slowly? Over smooth or bumpy ground, or through the air?
- **Geometry/Appearance** Does the target environment offer enough essential features (e.g. planes, corners, textures)?
- **Performance** Accuracy, computational efficiency, and how the approach scales with environment size.

Existing mapping systems tend to work well when operating under conditions similar to those in which their original authors demonstrated their performance. But in exploration tasks, our robots reliably discover vantage points with wildly varying feature availability while experiencing rapid, unexpected motions as they traverse unfamiliar terrain. Bumping and bouncing, slipping and sliding, or executing rapid pivots all have

The authors are with the GRASP Laboratory, School of Engineering and Applied Sciences, University of Pennsylvania, Philadelphia PA 19104.

Corresponding author: acowley@seas.upenn.edu

the potential to momentarily interfere with motion estimation by reducing overlap between consecutive sensor readings, or by breaking assumed motion priors used to improve optimization convergence (e.g. planar motion). The environment itself plays a role as transitions between enclosed and open spaces, or rounding a corner with poor visibility can lead to rapid significant changes in the types of visual and geometric features captured by a sensor. The challenges of these transient, yet frequently encountered, conditions may be diminished by ensuring that every scrap of information a sensor produces is leveraged by the SLAM system.

Consequently, we investigate a new SLAM method built to make use of as much range data as possible. The robustness of the approach is demonstrated along all three axes identified above by applying it to data collected by a sensor carried by a walking human, a running human, a small wheeled robot platform, a legged robot platform, an aerial robot platform, and an automobile, in environments from indoor clutter to underground tunnels to wild forests to public roads. Its performance is quantified, demonstrating trajectory error of 0.05% the trajectory length, its ability to run within a 15 W power budget, and its scaling from representing the geometry of window mullions to coping with 21.5 million cubic meters of space attached to a 4.4 km trajectory.

In this work, the environment is represented as a series of panoramic depth maps using a spherical projection, allowing us to represent each local map as a relatively small 2D array. Our implementation is able to effectively exploit the regular memory access patterns and parallelism that this representation affords. As a result it runs very quickly even on embedded GPU platforms. An important consequence is that we are able to make use of all available range measurements and avoid the need for early decisions about features. We argue that using all of the data provides additional robustness which allows the method to handle a wide variety of environments.

II. RELATED WORK

The SLAM problem has among the deepest wealth of research of any area of robotics. We draw particular inspiration from the idea of building an atlas [2] of local maps linked by rigid body transformations. As lidar sensors increase in resolution, keyframe techniques [3], [4] more commonly applied to camera-like sensors gain appeal. With such techniques, an instantaneous data capture serves as the initial value of a local map. In particular, we advocate the use panoramic image keyframes [5], [6] for their suitability to capture data gathered while

rotating. This paper improves on that described in [5] by showing how such panoramic representations can be efficiently constructed and registered online and in real-time.

Modern geometric SLAM systems, such as LOAM [7], have found success from winnowing the input data streams of high-resolution lidar through feature selection. In subsequent works, the authors improve robustness by adding complementary sensing modalities as fallbacks for when a primary sensor, typically a lidar, is unable to maintain continuous tracking [8]. LeGO-LOAM [9] is a notable specialization of LOAM to ground vehicle scenarios. It incorporates changes to point cloud filtering and feature selection that enable it to run at the sensor's update rate by tracking fewer features than the original LOAM formulation, but its requirement of a 9 degree-of-freedom IMU make it unsuitable for most of the experiments described here.

A parallel trend in mapping is driven by the great success found in the use of features detected by visible light and infrared sensitive cameras, with ORB-SLAM [10] a notable contributor to the popularization of monocular SLAM. It is able to extract and track features quickly enough for motion estimation, while the features it finds are also suitable for loop closing. However, a limitation of many texture-based feature mapping systems is that they produce a sparse 3D reconstruction of a workspace as defined by the tracked features. This limitation is somewhat alleviated by direct methods such as Direct Sparse Odometry [11] that optimize differences in image intensity values rather than extracted features. Such techniques can produce denser models than feature-based approaches since direct methods leverage more of the available image data. Work combining the dense 3D scanning of RGB-D sensors with more discriminative ORB features has been shown to work for moderately sized scenes[12].

Behley et al. [13] also argue for an approach to lidar mapping that make use of all available range measurements and describe how real-time SLAM can be performed on with an unordered surfel collection by projecting the surfels onto 2D vertex and normal maps. The approach described in this paper directly employs 2D array based representations for the local maps which can be efficiently updated in real-time even on embedded GPUs. While each panorama is only 2.5D, the *union* of these local maps at different vantage points allows us to effectively capture 3D space.

III. UPSLAM

Union of Panoramas SLAM (UPSLAM) represents a map as a graph whose nodes are panoramic depth images augmented with surface normal estimates, Figure 1, and whose edges are 3D transformations between those depth image keyframes. Keyframes may have different resolutions and extents than raw sensor data; we typically operate with a wider vertical field of view for keyframes

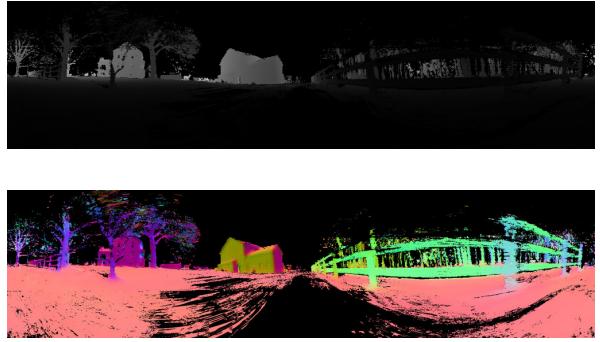


Fig. 1: Example keyframe panorama from a farm lane bordered by a fence. The upper image shows depth information, with lighter shades corresponding to greater depth, while the lower figure maps surface normal vector components to color.

than the sensor captures. The system keeps track of a current keyframe which is updated with each new lidar sweep. The experiments described here were conducted with an Ouster OS1-64 lidar which pairs an IMU providing accelerometer and gyroscope measurements at 100 Hz with an array of 64 lasers covering a 33° vertical field of view that are swept about an axis at 10 Hz.

A. Data Intake

During rapid motions, the 100 ms period of a lidar sweep may include significant sensor motions. Subsequent data processing will search for a rigid body transformation between 3D points derived from a single sweep and a model point set. The tension between applying a single transformation to points collected at different sensor poses is mitigated by using high-frequency inertial measurements to rotate all range measurements to the rotational frame of a single point in time [7]. The rotations based on IMU and lidar timing data are computed on the CPU, while virtually everything that follows is computed on the GPU.

The points resulting from this IMU motion correction are projected into a panoramic image. The image structure is helpful to compute a rough surface normal estimate as the cross product between the vectors defined by a central point and two of its image coordinate neighbors. These surface normal estimates, which are again represented as an image, are very noisy, but may be efficiently smoothed using an edge sensitive smoothing algorithm such as the à trous wavelet filter [14] to improve optimization convergence.

B. Motion Estimation

Each incoming panorama of depth and surface normal vectors is registered to the current keyframe using ICP with a point-to-plane error metric [15], [16], Equation 1.

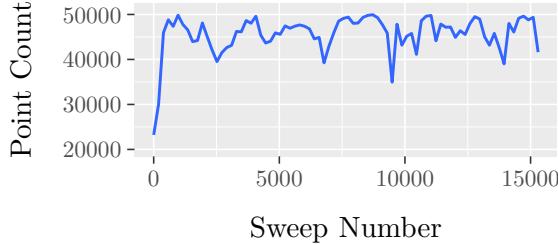


Fig. 2: Count of valid depth measurements over time for the Newer College Dataset.

$$E(T) = \sum_{(s_1, s_2) \in C \wedge \Omega_{d, \theta}(s_1, s_2)} |(Ts_2 - s_1) \cdot N(s_1)| \quad (1)$$

Here, we find a 3D transformation, T , that minimizes the distance between each pair of points drawn from the set of projective correspondences, C , when projecting the vector between corresponding points onto the normal vector of the first, $N(s_1)$. A similarity filter, Ω , is applied to the set of correspondences to reject those for which the distance, d , between the two points is greater than some threshold, or when the angle, θ , between their associated surface normal vector estimates is too great. This method may be efficiently performed with a GPU [17].

A critical factor in the robustness of the mapping system is how many points are incorporated into this optimization. Figure 2 shows the number of valid points over time for a typical experiment, where our notion of validity is defined as a point deriving from a valid range estimate (the sensor will not return a measurement for every pixel in every sweep), and an intensity above a small threshold (this rejects points whose range estimate is based on very little received light). In the example shown in the figure, the sensor starts out near a wall in a tunnel, but once it emerges into the open, 40 to 50 thousand points are used at the 10Hz sweep rate.

C. Keyframe Update

The optimized transformation provides a basis for a final set of projective correspondences, in which each keyframe pixel is projected onto the sweep depth image. Just as with the ICP optimization, a similarity filter is used to determine if the new sweep data should be averaged in to the keyframe pixel. A weighted average is used when updating the keyframe: each keyframe pixel tracks how many samples have been incorporated into it, and that number – saturated at 10 in our system – is used to determine the weightings of the averaging procedure. The simplicity of this averaging scheme is offset by the use of multiple keyframes in the map representation when considered holistically. Moving objects can strain the balance of smoothing and responsiveness within a single keyframe, but considering multiple keyframes reconstructs a latent multi-hypothesis model of 3D occupancy which is an effective tool for filtering moving objects

from the map. This mutual consistency check between keyframes is performed whenever switching keyframes in order to reduce the impact of moving objects.

Once updated, a decision is made as to whether we should continue using the current keyframe. If the registration quality – defined as the ratio of ICP registration inliers to valid depth measurements in the sweep – has dropped below a configurable threshold ¹, a new keyframe is sought. The pose graph is first consulted to identify the nearest keyframes to the current pose estimate to detect potential loop closures.

Since these loop closure registrations are likely to occur under the influence of significant drift, a grid search is undertaken in which transformations are evaluated with $\frac{1}{8}$ -resolution projective alignment checks. The best candidate is promoted to a full-resolution ICP registration. If the registration quality is higher than the threshold for creating a new keyframe, the pose graph is updated with an edge between the current keyframe and the neighbor, and then that neighbor becomes the new current keyframe. Otherwise, if that ratio is higher than 75% of the new keyframe inlier ratio, the pose graph is updated with the transformation computed from the registration, but the current keyframe is not updated. We refer to these as *strong* and *weak* loop closures: both varieties improve the global accuracy of the map, but strong loop closures additionally reduce the number of keyframes created when a location is re-visited.

The pose graph structure is reflected in GTSAM [18], which is used to globally optimize inter-keyframe transformations on the CPU. Finally, a complementary filter [19] on the IMU data is used to maintain a smoothed estimate of Earth's gravity vector which is attached to each keyframe as it is added to GTSAM. The observations of this shared phenomenon help minimize the effects of errors in pitch, which are relatively common given the limited vertical field of view of the sensor [20].

IV. QUALITATIVE BREADTH

A. University - Indoor

A first test of the mapping software is a 145 m walk around a lab space, moving at speed of 1–2 m/s. The map shown in Figure 3 was captured with 50 keyframe images occupying 80 MiB on disk. While the motion here is benign, the abundance of glass around the angular section at the bottom of the figure, along with a few narrow passages, provides some challenge due to the sensor's inability to reliably detect glass or walls closer than around 50 cm.

B. University - Outdoor

The sensor was then taken outside, carried overhead, and run along a 375 m loop around a cluster of buildings on a University campus at 2–3 m/s. The resulting point cloud, shown in Figure 4, benefits from the single

¹This configuration parameter was allowed to vary between experiments.

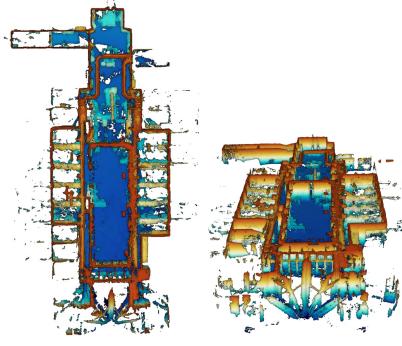


Fig. 3: A 145 m trajectory through a cluttered lab space. The ceiling was removed for visualization clarity.

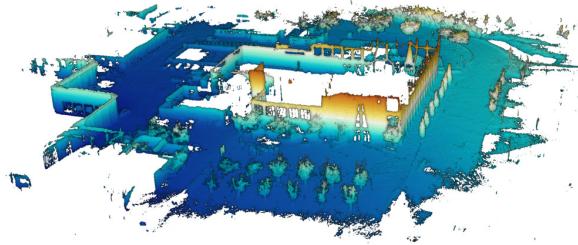


Fig. 4: A running loop around the Pennovation campus with the sensor carried overhead.

loop closure formed when the sensor came within approximately 10 m of its starting location . The running pace was chosen to stress non-planar motion with high amplitude rotations due to shaking and bouncing.

C. Farm

A Clearpath Robotics Jackal, Figure 5a, was driven along a 731 m route along a driveway, dirt trail, and grass on a farm, Figure 5b. This excursion presented several challenges due to the rough terrain and varied surroundings. Tall vegetation waved in the breeze, while the pond itself offered minimal range measurements. Data collection included intermittent occlusion of the sensor, along with sharp accelerations to get the vehicle un-stuck.

D. Underground

A Ghost Robotics Vision 60 quadruped robot was walked along a 220m trajectory in an old coal mine, capturing the map shown in Figure 6 with just 44 panoramic images. This experiment provided a substantially different motion model than those previously described as the wet gravel floor of the mine meant that the robot had to continually work to maintain its balance. Dripping water, mist, and dust contribute a substantial amount of noise to individual lidar sweeps. Additionally, the tunnels visited ranged from wide access to the main elevator, bottom-center of Figure 6c, to a narrow, angular tunnel

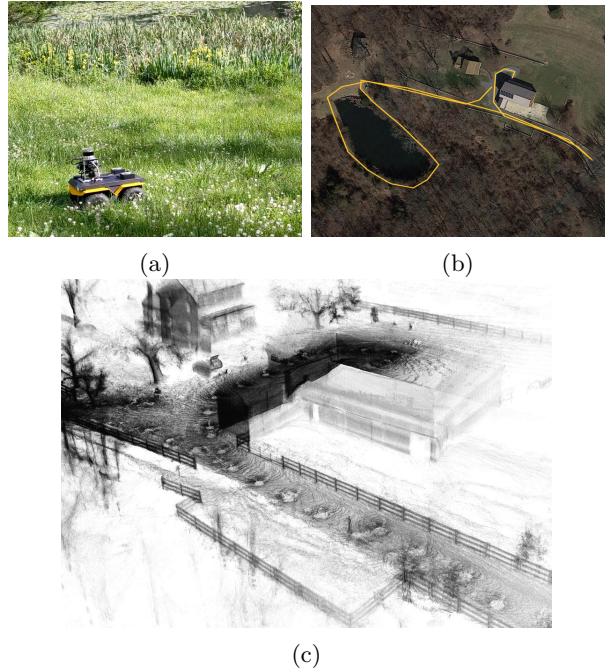


Fig. 5: The Miller family farm in central Pennsylvania (b). The 731m on- and off-road trajectory around pond is captured with geometric detail including individual rails on gates and fences. (c).

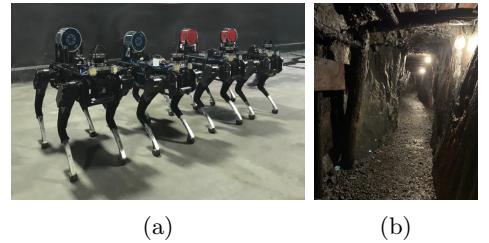


Fig. 6: A quadrupedal robot (a) was walked around a coal mine (b) along a 220m trajectory. The resulting map (c) was captured with 44 images.

approximately 1 m across, Figure 6b, visible on the right side of the point cloud.

E. Aerial

The Wharton State Forest in New Jersey is an expansive wilderness with dense tree cover. Flying a quadrotor², Figure 7a, in this environment is challenging for

²Data courtesy of Treeswift.

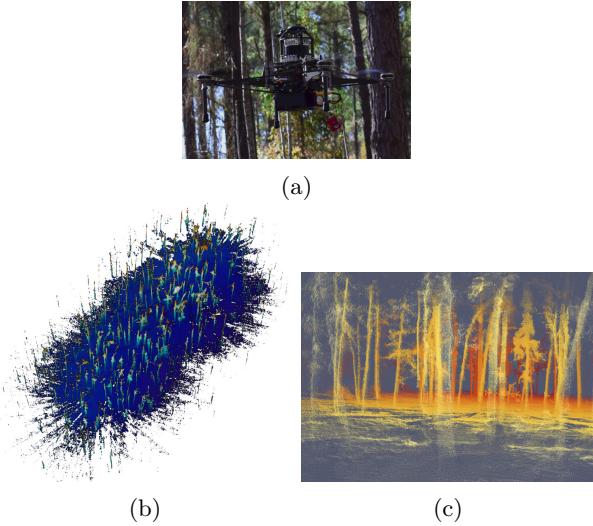


Fig. 7: A quadrotor was flown along a 117 m trajectory in a dense forest. The entire environment (b) is devoid of manmade structure, but the trunks and branches of the many trees are cleanly captured in the point cloud (c).

many reasons, among which are the challenges of building a map in a region with no large planar surfaces and complex occlusions. UPSLAM’s use of dense surface measurements extracts as much as possible from the sensor, and it is able to track its motion along a 117 m trajectory while producing a map populated with crisply defined trees, Figure 7. This map is captured in 30 keyframe images occupying 40 MiB on disk, sufficient to produce the point cloud shown in Figure 7c.

F. Automotive

While each of the previous experiments subjected the system to different motion models, the fastest the sensor was moved was about 3 m/s when carried by a running human. This section presents a dramatically different scenario: the sensor was mounted atop an automobile and driven 4.4 km through a town at up to 50 km/h (31 miles per hour), Figure 8.

This fast-moving data set³ prompted 70 loop closures, and produced 1045 keyframe images occupying 1220MiB on disk. Map construction involved 20GiB of raw, uncompressed keyframe data at runtime, however UPSLAM is able to stream keyframes between disk and GPU memory as needed, allowing us to cap the resident set at 3GiB of GPU memory. Because the sensor was moving so quickly – up to 1.4 m per sweep – there is far less redundancy in the collected lidar point cloud, limiting opportunities for compression, moving object filtering, or even surface smoothing. However, considerable structural detail is still apparent despite the presence of other vehicles on the road.

³A second configuration parameter was used for this experiment to enable a zero acceleration motion prior, as opposed to the default zero velocity prior used in all other experiments which involve more sudden reversals of velocity.

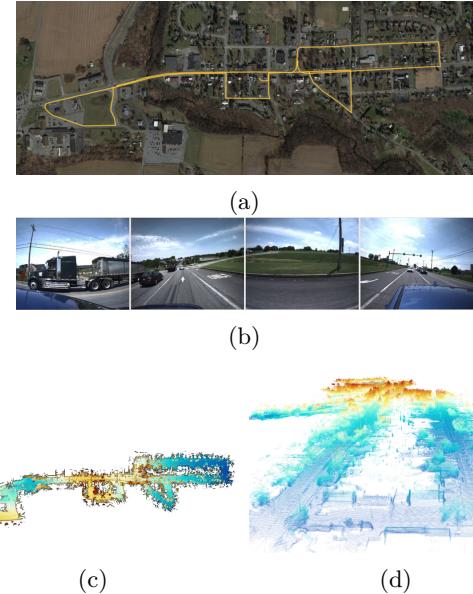


Fig. 8: A 4.4km route driven through Morgantown, Pennsylvania, highlighted in Google Earth (a), seen from street level (b), and viewed as a point cloud produced by UPSLAM (c). A lower view, (d), shows structural detail picked up by the single sensor moving at road speeds.

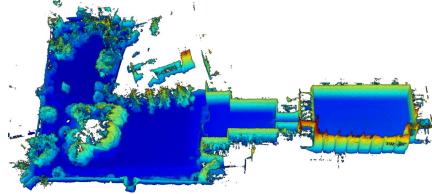


Fig. 9: An overhead view of a point cloud representing New College, Oxford University .

V. QUANTITATIVE DEPTH

A. Newer College

The Newer College Dataset [21] provides an opportunity to quantify trajectory accuracy. Point clouds captured by an Ouster OS1-64 were registered against a point cloud collected by a tripod-mounted, survey grade Leica BLK360 scanner. A sense of the space may be found in the point cloud rendering shown in Figure 9, showing the quad on the lower right, the midsection, and then the park on the left. The sensor was carried on a circuitous route making two full laps around the college grounds⁴.

1) *Trajectory Accuracy:* Using time stamps to align the UPSLAM and ground truth trajectories shows broad agreement, Figure 10, with most of the discrepancy introduced around a particular corner of the park area visible at the bottom-right of the figure. The temporally aligned trajectories provide a set of point correspondences. We use this correspondence set to compute a 3D transforma-

⁴This is the short_experiment of the dataset.

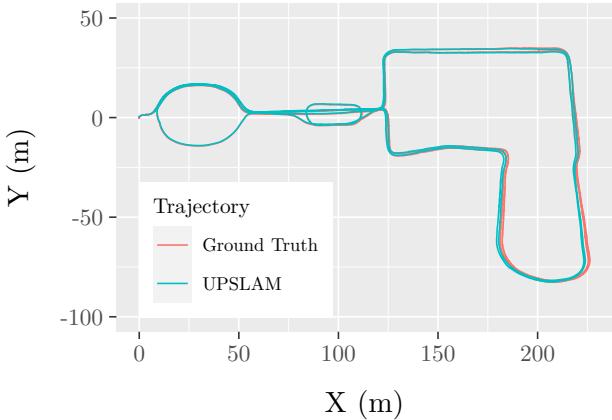


Fig. 10: Comparison of ground truth and UPSLAM trajectories for the Newer College Dataset.

Mode	CPU Cores	CPU MHz	GPU MHz	Memory MHz
Max	8	2265.6	1377	2133
30W	6	1450	900	1600
15W	4	1200	670	1333

TABLE I: Jetson Xavier Power Modes

tion that optimally aligns the trajectories with respect to squared Euclidean distance between each member of each correspondence. The correspondences aligned in this way have a median distance of 0.62 m along the entire 1.4 km trajectory. This metric is similar to the Absolute Trajectory Error (ATE) [22], and is designed to show global consistency between trajectories. The ATE uses the root mean square error rather than the median, which results in an ATE of 0.77 m for this trajectory alignment, or 0.05% of the total trajectory length.

2) *Computational Performance:* The NVIDIA Jetson AGX Xavier embedded platform may be configured to run with various limits on its power consumption. These limits are observed by lowering the clock frequencies of the CPU and GPU, and disabling some of the eight CPU cores. The highest power mode runs at the maximal clock frequencies, and is limited only by the 65W power adapter. In this configuration, we consider high, 2048x256, and low, 1024x128, resolution keyframes, Figure 11. The next power limits we consider are 30W and 15W, detailed in Table I⁵.

The time taken to incorporate a new lidar sweep into the map is shown for each configuration in Figure 11. The median time taken with the Xavier’s maximal power budget and high resolution keyframes is 6 ms. The figure shows a kernel density estimate of update times taken while processing the Newer College Dataset to reflect the fact that not every update takes the same amount of time: some updates trigger new keyframe creation, and some may result in relatively slow graph updates due

⁵The 15 and 30W power levels tested are models 2 and 4 of NVIDIA’s nvpmodel utility for the Xavier AGX.

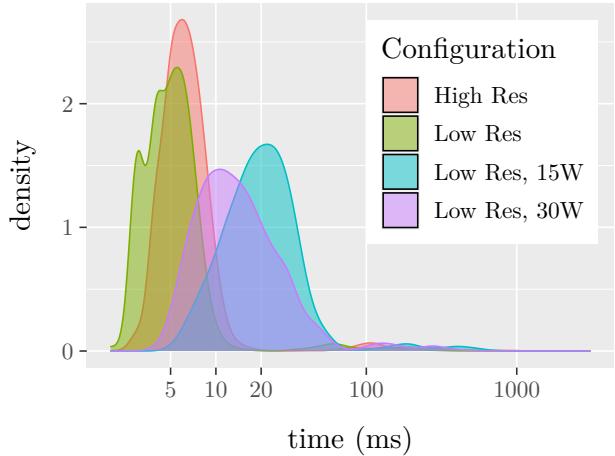


Fig. 11: Kernel density estimate of time to compute SLAM updates on an NVIDIA Jetson AGX Xavier at different keyframe resolutions and computer power limits.

to loop closures. This long tail of update times is more apparent when considering low resolution keyframes with maximal, 30W, and 15W power budgets, which resulted in median update times of 5 ms, 13 ms, and 20 ms. Inputs are buffered when graph optimization takes longer than the interval between sweeps. The Xavier’s 512 CUDA cores running at full power with this workload exhibit similar performance to NVIDIA’s Quadro P1000 or GTX 1050 GPUs.

VI. CONCLUSIONS

The UPSLAM mapping system has been qualitatively evaluated across diverse environments spanning indoor, outdoor, underground, aerial, and automotive settings. These challenging scenarios involving mud, robot collisions, dense foliage, bodies of water, and road traffic were all handled by the same mapping software adjusted with only two configuration parameters. Beyond qualitative checks for functionality, accuracy was quantified with an absolute trajectory error of less than 0.1% of the total trajectory length. Performance was measured at a median update rate of less than 10 ms on an embedded compute platform. Workspace scaling has been demonstrated by navigating corridors and tunnels 1 m across, as well as multi-kilometer road ways. We believe that this effort pushes the boundaries of robustness and performance for SLAM software, while demonstrating the effectiveness of the graph of panoramas map representation and the design choice to use all available range data during registration.

REFERENCES

- [1] 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, pp. 1309–1332, 2016.

- [2] M. Bosse, P. Newman, J. Leonard, and S. Teller, "An atlas framework for scalable mapping," in *in IEEE International Conference on Robotics and Automation*, 2003, pp. 1899–1906.
- [3] M. Meilland and A. Comport, "On unifying key-frame and voxel-based dense visual SLAM at large scales," in *International Conference on Intelligent Robots and Systems*. Tokyo, Japan: IEEE/RSJ, 3-8 November 2013.
- [4] M. Meilland, A. I. Comport, and P. Rives, "Dense omnidirectional RGB-D mapping of large scale outdoor environments for real-time localisation and autonomous navigation," *Journal of Field Robotics*, vol. 32, no. 4, pp. 474–503, June 2015.
- [5] C. J. Taylor, A. Cowley, R. Kettler, K. Ninomiya, M. Gupta, and B. Niu, "Mapping with depth panoramas," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Sept 2015, pp. 6265–6272.
- [6] A. Milioto, J. Behley, C. McCool, and C. Stachniss, "Lidar panoptic segmentation for autonomous driving," in *IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS*, 10 2020.
- [7] J. Zhang and S. Singh, "LOAM: Lidar odometry and mapping in real-time," in *Robotics: Science and Systems Conference (RSS)*, July 2014.
- [8] ———, "Laser-visual-inertial odometry and mapping with high robustness and low drift," *Journal of Field Robotics*, vol. 35, no. 8, pp. 1242–1264, 2018.
- [9] T. Shan and B. Englot, "Lego-loam: Lightweight and ground-optimized lidar odometry and mapping on variable terrain," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 4758–4765.
- [10] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardós, "Orbslam: A versatile and accurate monocular slam system," *IEEE Transactions on Robotics*, vol. 31, no. 5, pp. 1147–1163, Oct 2015.
- [11] J. Engel, V. Koltun, and D. Cremers, "Direct sparse odometry," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, mar 2018.
- [12] H. Liu, C. Li, G. Chen, G. Zhang, M. Kaess, and H. Bao, "Robust keyframe-based dense SLAM with an RGB-D camera," *CorR*, vol. abs/1711.05166, 2017.
- [13] J. Behley and C. Stachniss, "Efficient surfel-based slam using 3d laser range data in urban environments," in *RSS*, 06 2018.
- [14] H. Dammertz, D. Sewitz, J. Hanika, and H. Lensch, "Edge-avoiding À-trous wavelet transform for fast global illumination filtering," vol. 2010, 01 2010, pp. 67–75.
- [15] P. Besl and N. D. McKay, "A method for registration of 3-d shapes," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 14, no. 2, pp. 239–256, 1992.
- [16] Y. Chen and G. Medioni, "Object modelling by registration of multiple range images," *Image and Vision Computing*, vol. 10, no. 3, pp. 145–155, Apr. 1992.
- [17] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohli, J. Shotton, S. Hodges, and A. Fitzgibbon, "Kinectfusion: Real-time dense surface mapping and tracking," in *Proceedings of the 2011 10th IEEE International Symposium on Mixed and Augmented Reality*, ser. ISMAR '11. Washington, DC, USA: IEEE Computer Society, 2011, pp. 127–136.
- [18] F. Dellaert, "Factor graphs and GTSAM: A hands-on introduction," 2012.
- [19] R. G. Valenti, I. Dryanovski, and J. Xiao, "Keeping a good attitude: A quaternion-based orientation filter for imus and margs," *Sensors*, vol. 15, no. 8, pp. 19 302–19 330, 2015.
- [20] P. Puri, D. Jia, and M. Kaess, "GravityFusion: Real-time dense mapping without pose graph using deformation and orientation," in *IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS*, Vancouver, Canada, Sept. 2017, pp. 6506–6513.
- [21] M. Ramezani, Y. Wang, M. Camurri, D. Wisth, M. Mattamala, and M. Fallon, "The newer college dataset: Handheld lidar, inertial and vision with ground truth," 2020.
- [22] 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*, 10 2012, p. nil.