2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2025)
October 19-25, 2025. Hangzhou, China

2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) | 979-8-3315-4393-8/25/$31.00 ©2025 IEEE | DOI: 10.1109/IROS60139.2025.11245902

M2P2: A Multi-Modal Passive Perception Dataset for Off-Road Mobility in Extreme Low-Light Conditions

Aniket Datar*[1], Anuj Pokhrel*[1], Mohammad Nazeri*[1], Madhan B. Rao*[1], Harsh Rangwala[1], Chenhui Pan[1], Yufan Zhang[1], André Harrison[2], Maggie Wigness[2], Philip R. Osteen[2], Jinwei Ye[1], and Xuesu Xiao[1]

***Abstract*— Long-duration, off-road, autonomous missions require robots to continuously perceive their surroundings regardless of the ambient lighting conditions. Most existing autonomy systems heavily rely on active sensing, e.g., LiDAR, RADAR, and Time-of-Flight sensors, or use (stereo) visible light imaging sensors, e.g., color cameras, to perceive environment geometry and semantics. In scenarios where fully passive perception is required and lighting conditions are degraded to an extent that visible light cameras fail to perceive, most downstream mobility tasks such as obstacle avoidance become impossible. To address such a challenge, this paper presents a Multi-Modal Passive Perception dataset, M2P2, to enable off-road mobility in low-light to no-light conditions. We design a multi-modal sensor suite including thermal, event, and stereo RGB cameras, GPS, two Inertia Measurement Units (IMUs), as well as a high-resolution LiDAR for ground truth, with a multi-sensor calibration procedure that can efficiently transform multi-modal perceptual streams into a common coordinate system. Our 10-hour, 32 km dataset also includes mobility data such as robot odometry and actions and covers well-lit, low-light, and no-light conditions, along with paved, on-trail, and off-trail terrain. Our results demonstrate that off-road mobility and scene understanding under degraded visual environments is possible through only passive perception in extreme low-light conditions. The project website can be found at `https://cs.gmu.edu/~xiao/Research/M2P2/`.**

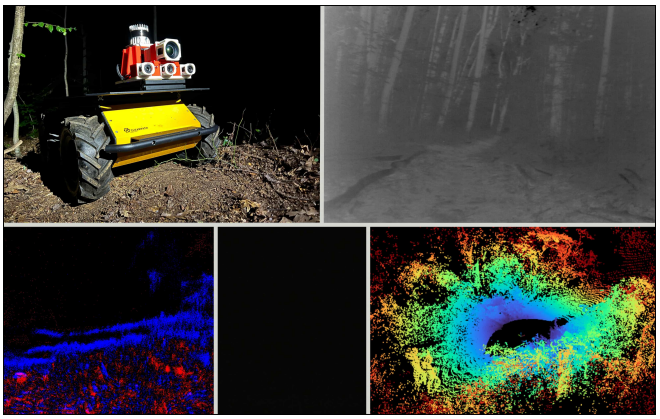


Fig. 1: Multi-Modal Passive Perception Data Collection in an Off-Road Forest Environment in Complete Darkness. Top Left: Clearpath Husky with the Sensor Suite (flashlight for visualization only); Top Right: Thermal Image; Bottom Left: Event Stream; Bottom Middle: RGB Image (fail to perceive); Bottom Right: LiDAR Point Cloud (for ground truth).

I. Introduction

Autonomous mobile robots have found their way out of controlled lab, factory, and warehouse environments into the wild [1]. On their way to deliver packages [2], inspect infrastructure [3], maintain agricultural fields [4], and conduct search and rescue missions [5], those robots constantly perceive their surroundings with their onboard sensors. The perceived geometric and semantic world representations allow them to move to their goals while avoiding collisions. Such an extension in Operational Design Domain requires robot perception systems to address challenges around the clock, ranging from well-lit to no-light conditions, as well as from paved to completely off-road terrain in the wild.

Existing perception systems for mobile robots rely heavily on active sensing. For example, LiDAR range finders [6] use pulsed laser beams to detect distance and perceive environmental geometry, while Time-of-Flight sensors [7] use infrared light and measure the time it takes for the light signal to travel to the target and back. Despite working well in all lighting conditions, many active sensors suffer from significant noise in heavy rain, snow, and fog. Furthermore, the reliance on the emission of active light signals will expose the presence of the robot, making those active sensors less ideal for covert operations, e.g., in military settings.

Non-active, visible light imaging sensors, e.g., RGB cameras, are also widely used in robot perception systems, relying on reflected light to form images for non-light emitting objects. Stereo camera pairs can triangulate to determine distance and use different RGB color channels to reason about semantics. Those sensors work well in well-lit indoor and outdoor environments and provide similar sensing as human perception. However, visible light imaging sensors require good lighting conditions to perceive reflected light and form visible pixels, and therefore suffer from degraded perception quality in low-light to no-light conditions.

These aforementioned limitations of existing active and visible light imaging sensors present challenges for long-

[1]George Mason University {adatar, apokhre, mnazerir, mbalajir, hrangwa, cpan7, yzhang82, jinweiye, xiao}@gmu.edu
[2]DEVCOM Army Research Laboratory {andre.v.harrison2.civ, maggie.b.wigness.civ, philip.r.osteen.civ}@army.mil
*Equally contributing authors

This work has taken place in the RobotiXX Laboratory at George Mason University. RobotiXX research is supported by National Science Foundation (NSF, 2350352), Army Research Office (ARO, W911NF2320004, W911NF2420027, W911NF2520011), Air Force Research Laboratory (AFRL), US Air Forces Central (AFCENT), Google DeepMind (GDM), Clearpath Robotics, Raytheon Technologies (RTX), Tangenta, Mason Innovation Exchange (MIX), and Walmart.

Please refer to https://arxiv.org/abs/2410.01105 for the full version.

duration, off-road, autonomous missions, since robots need to perceive their surroundings around the clock regardless of the ambient lighting conditions and are also oftentimes required to be fully passive to maintain stealth. To operate in low-light to no-light conditions without emitting any active light signatures, novel sensing modalities, including thermal and event cameras, show promise by passively sensing infrared radiation from all objects with a temperature above absolute zero or per-pixel brightness changes (also called "events") asynchronously with low latency, high dynamic range, and low power consumption, respectively.

In this paper, we propose to use multi-modal passive perception modalities to enable robot perception in extreme low-light conditions so as to facilitate downstream off-road mobility tasks (Fig. 1). To be specific, our contributions include:

- a multi-modal sensor suite including thermal, event, and stereo RGB cameras, GPS, two IMUs, and a high-resolution LiDAR for ground truth;
- a precise multi-sensor calibration procedure for multi-modal perceptual streams;
- a Multi-Modal Passive Perception dataset, M2P2, with data ranging from different lighting conditions (well-lit to no-light) and various off-road terrain conditions (paved to off-trail), along with mobility data like robot odometry and actions; and
- experimental results demonstrating off-road mobility, depth reconstruction, and vehicle odometry through only passive perception in extreme low-light conditions.

II. Related Work

In this section, we review related work in off-road perception systems and passive perception sensors.

A. Off-Road Perception

Perception in off-road environments requires both exteroceptive and interoceptive sensing to understand the environment and the robot's interaction with it. The availability of a wide array of sensors makes safe traversal through off-road environments possible. While a single modality may suffice for navigation in structured environments, the inclusion of multiple modalities in challenging environments adds robustness and redundancy, ensuring that navigation can continue even if one or more sensors are unable to work at full capacity because of adverse environmental conditions. By combining complementary data from multiple sensors, robots can also better perceive and interpret complex environmental features for comprehensive understanding in a variety of off-road unstructured scenarios.

Active sensing modalities like LiDAR and RADAR detect and perceive environmental geometry, enabling the creation of 2D, 3D, or 2.5D elevation maps [11]–[15] of the environment. Although LiDAR-based systems are highly popular for their robustness and precision, they can suffer in heavy rain, snow, and fog, and may struggle to map terrain at greater distances [16]. Additionally, the use of pulsated beams can expose the presence of the robot. On the other hand, vision-based navigation systems utilize visible light imaging sensors, e.g., RGB or RGB-D cameras, to understand the terrain semantics [17]–[20], create elevation maps [16], [18], and map off-road terrain [21], [22]. Although vision-based navigation systems are advantageous due to their passive sensing capabilities and ability to provide rich environmental information, their reliance on visible light causes poor performance in low-light conditions. While also being passive, interoceptive sensors like IMUs and force sensors measure robot internal states during environment interactions, which can be used to generate traversability maps [17], [23] and model terrain response [24], [25] when combined with exteroception.

Combining the advantages of the aforementioned perception modalities expands robots' Operational Design Domain in varying environmental conditions around the clock, such as low visibility or extreme weather, with the possibility of staying passive. With the recent advancement in data-driven approaches [1], multi-modal off-road datasets [8], [26], [27] are essential for developing and refining perception and mobility algorithms, providing a foundation for training, testing, and benchmarking. Our multi-modal sensor suite offers passive sensing capabilities with precise ground truth from active perception, enabling navigation in extremely low-light off-road environments. The sensor suite is resilient to environmental degradation like dust, smoke, fog, snow, and rain, and can be calibrated in a single step for effective off-road navigation.

B. Related Datasets

A few existing datasets provide a variety of sensor modalities and ground truth data, enabling the development and benchmarking of algorithms in areas such as SLAM, object recognition, and autonomous navigation (Table I): MVSEC [28] is the first dataset that synchronizes stereo event cameras and provides accurate ground truth depth from LiDAR and SLAM and ground truth pose using a motion capture system and GPS; UZH-FPV [29] dataset utilized fast, aggressive, and agile drones to capture event camera data for extreme motion scenarios, but does not contain depth information; For night and day place recognition tasks, Maddern and Vidas [30] built a capture platform consisting of GPS, RGB camera, and thermal camera to capture data from before dawn to after dusk; The KAIST Multi-Spectral Day/Night Dataset [31] introduced a sensor system designed for SLAM, comprising stereo RGB cameras, LiDAR, and thermal camera; Aiming at off-road environments such as forests and urban areas, M3ED [32] used high resolution stereo event cameras, grayscale and RGB cameras, IMU, LiDAR, and RTK localization to collect a high-speed dynamic motion dataset; ViViD++ [8] is the first dataset to feature aligned information from multiple types of alternative vision sensors, including RGB, thermal, event, depth, and inertial measurements. Compared to existing datasets, our M2P2 dataset is the first dataset that focuses on off-road mobility in extremely low-light environments with the most perception modalities and highest sensor quality, as well as

TABLE I: Comparison with alternative vision datasets.

Dataset	Sensor Modality							Hardware	Environments	Lighting
	RGB	Depth	Thermal	Event	LiDAR	IMU	GPS			
ViViD++ [8]	✓	✓	✓	✓	✓	✓	✓	Vehicle	Indoor/Urban	Day/Night
DiTer++ [9]	✓	✓	✓	✗	✓	✓	✓	Legged	Diverse Terrain	Day/Night
TartanDrive 2.0 [10]	✓	✓	✗	✗	✓	✓	✓	Wheeled	Off-road	Day
M2P2	✓	✓	✓	✓	✓	✓	✓	Wheeled	Off-road	Day/Night

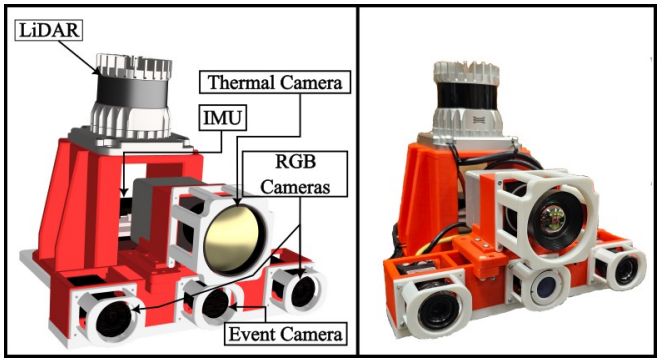


Fig. 2: Sensor Suite CAD (Left) and Hardware (Right).

a precise multi-modal calibration procedure with accurate synchronization (see Table I for comparison).

III. Multi-Modal Sensor Suite

Our multi-modal sensor suite comprises a thermal and an event camera, stereo RGB cameras, two IMUs, GPS, and LiDAR for ground truth. All sensors are assembled in a custom-designed 3D-printed structure, which can be easily mounted on most mobile robot platforms (Fig. 2). The total dimensions of the sensor suite are 0.31×0.26×0.24 m, with a total weight of 2 kg.

A. Thermal Camera

Our sensor suite includes the Xenics Ceres T 1280 thermal camera, offering high-resolution LWIR imaging at 1280×1024. It captures up to 45 FPS over a GigE Vision interface. Paired with an 11 mm wide-angle lens (71.7° HFoV, 58.9°, f/1.2 aperture), it delivers exceptional thermal image quality surpassing that of existing open-source datasets.

B. Event Camera

We use the Prophesee Metavision EVK4 event camera with a 1280×720 resolution and 220μs latency in a compact form. It achieves a time resolution of 10K FPS and operates in low light down to 0.08 lx. A 46.8° HFoV, 36° VFoV lens (aperture f/2–11, fixed at f/4.0) is used. To suppress LiDAR-induced noise, we place an IR filter over the lens.

C. Stereo RGB Cameras

We use two FLIR Blackfly S cameras for RGB imaging at a resolution of 1616×1240, captured at up to 175 FPS (fixed at 10 FPS). While stereo RGB fails in complete darkness, it remains effective in partially degraded or low ambient light environments.

D. IMUs

We use a Yahboom 10-DoF IMU featuring a 3-axis accelerometer, 3-axis gyroscope, 3-axis magnetometer, and a barometer. The sample rate of the IMU is 200 Hz. It features built-in data fusion and gyro stabilization. We also include the IMU embedded in the LiDAR (see details below).

E. LiDAR for Ground Truth

A 3D Ouster OS1-128 LiDAR is used to provide ground truth with 128 lines of vertical divisions in 45° VFoV and selectable 512, 1024, and 2048 angle divisions in 360° HFoV at 10/20 Hz. For best data efficiency, LiDAR point clouds are recorded with 1024 angle divisions at 10 Hz. The LiDAR also features a built-in 6-DoF IMU with a 125 Hz sample rate for LiDAR frame calibration.

IV. Sensor Suite Calibration

To interpret how real-world features in world coordinates map to sensor readings and relate across modalities, we develop a streamlined calibration procedure to align all sensors in the suite within a common coordinate frame.

Traditional methods rely on standard printed geometric targets like checkerboards for cameras or flat surfaces for LiDAR to estimate intrinsics and extrinsics. However, these are ineffective for our multi-modal setup: thermal cameras can't see standard targets in IR, and event cameras require motion to detect intensity changes. Thus, we need a common calibration target perceivable by all sensors to calibrate both intrinsic and extrinsic parameters.

A. Thermal Checkerboard

The first calibration challenge comes from the thermal camera, which needs varying thermal signatures to detect geometric features. To create a thermal contrast, we build a calibration target from a 3 mm thick aluminum sheet with attached carbon fiber squares of 35 mm length, CNC-milled to 0.05 mm accuracy. As aluminum reflects strongly in the long-wave IR spectrum (like a mirror in visible light), we anodize it to reduce unwanted reflections. Heating the target to 45°C reveals the checkerboard pattern in thermal images due to the large emissivity difference between aluminum and carbon fiber. (Fig. 3 left). Due to the contrast in color of aluminum and carbon fiber, the same pattern is visible in both RGB cameras (Fig. 3 right).

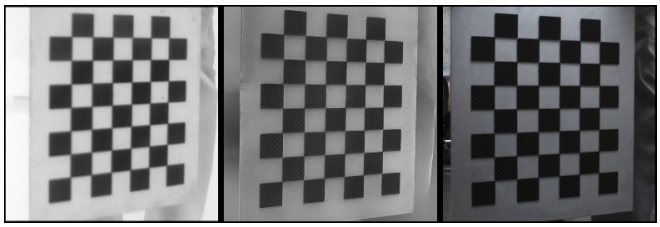


Fig. 3: Calibration Target (Thermal, Event, and RGB Image).

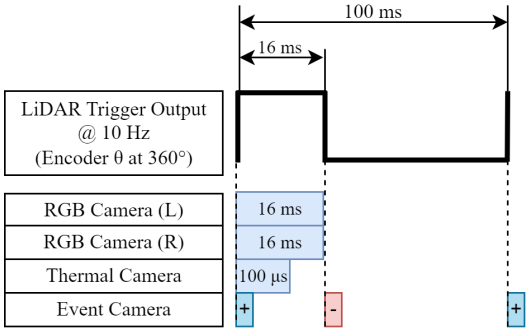


Fig. 4: Multi-Modal Synchronization: LiDAR trigger synchronized to internal encoder angle ($\theta = 360°$) initiates frame acquisition at a rate of 10 Hz for RGB and thermal cameras, with event camera recording trigger edges for frame reconstruction.

B. Event Reconstruction

To address the second challenge of aligning asynchronous event data with synchronous streams like thermal and RGB images, we use a two-step approach. First, we reconstruct grayscale images from the raw event stream using E2Calib [33] (Fig. 3 middle). We also use the event camera's trigger input to mark precise timestamps for frame reconstruction, enabling accurate temporal alignment with other sensors. This approach overcomes the asynchronous nature of event data and builds a reliable temporal link with synchronous streams, supporting multi-modal fusion and calibration.

C. Multi-Modal Synchronization

With a shared calibration target visible to all four cameras, including the second RGB camera in the stereo pair, the final challenge is synchronizing multiple asynchronous data streams to achieve calibration convergence. To solve this, we implement a synchronization scheme shown in Fig. 4. All cameras are synchronized to the LiDAR, which emits a sync pulse at 10 Hz aligned to its encoder angle at 360°. This pulse triggers frame capture in the RGB and thermal cameras and timestamps events in the event camera. The pulse width matches the RGB exposure time, and the falling edge defines the event camera's frame boundary, aligning it with RGB frame completion and ensuring temporal correlation across all sensors.

D. All-in-One Calibration Procedure

Finally, we combine all synchronized frames into a ROS bag, which is compatible with standard calibration toolkits.

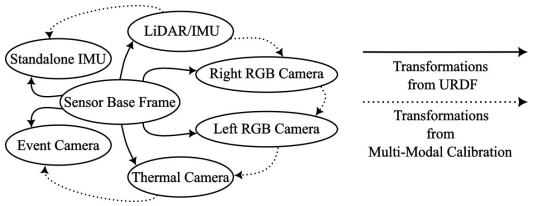


Fig. 5: Transformation Tree of the Sensor Suite: Solid arrows indicate direct hardware transformations, while dotted arrows represent transformations from our multi-modal calibration.

Fig. 6: Multi-modal data from the M2P2 dataset, showcasing spatial and temporal alignment in a low-light, off-road forest environment. LiDAR point cloud overlaid on RGB image (left), reconstructed event frame at the trigger's falling edge (middle), and thermal image (right).

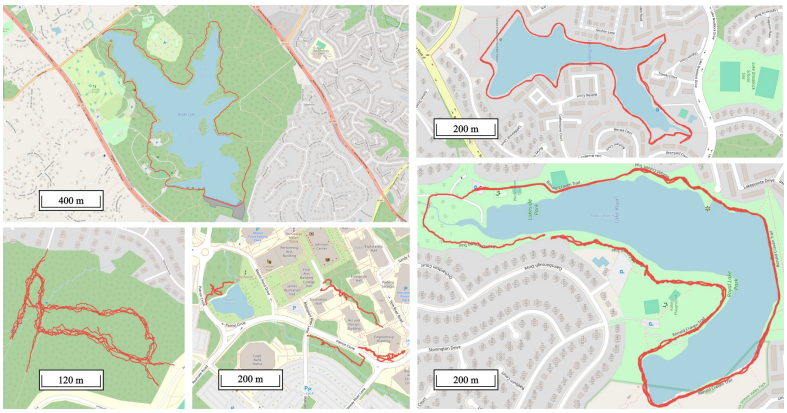
In our implementation, We use Kalibr [34] to estimate camera intrinsics and extrinsics, and further calibrate cameras to IMUs to complete the sensor suite's transformation tree. Since the Ouster IMU includes a factory-calibrated 6-DoF transform to the LiDAR base, we use the LiDAR base as the reference frame to unify all sensors. The full transformation tree from both calibration and hardware design is shown in Fig. 5.

Fig. 6 shows the LiDAR point cloud overlaid on the corresponding RGB image, along with the reconstructed event frame and thermal image, demonstrating the spatial and temporal alignment of the multi-modal data.

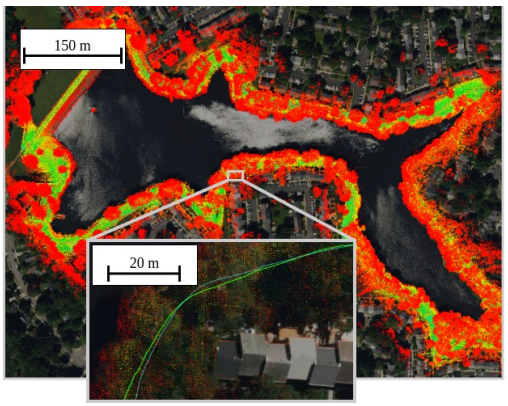
V. Multi-Modal Passive Perception Dataset

M2P2 dataset contains over 10 hours of data collected across diverse and challenging terrain conditions (Fig. 7a). Data were recorded using the sensor suite mounted on a Clearpath Husky A200 robot. Sequences span a wide range of environments—from paved trails to unpaved paths and unprepared off-trail areas in dense forests with thick vegetation and narrow passages. To capture varied lighting, data were collected at dusk with illumination levels ranging from 20 lx to complete darkness (0 lx). This ensures the dataset supports navigation tasks in both well-lit and low-visibility conditions.

The dataset is organized as ROS-bag files and includes compressed RGB and thermal images at 10 FPS, asynchronous raw event data, LiDAR point clouds, IMU readings, GPS coordinates, robot odometry and status, and joystick commands. All camera data are synchronized using LiDAR trigger pulses to ensure temporal alignment across modalities. Due to dense forest canopy, GPS is available for only 87.97% of the dataset. However, LiDAR, IMU, and GPS (when present) can be fused using LIO-SAM, which primarily relies on lidar-inertial odometry. Fig. 7b shows a LIO-SAM generated map overlaid on a satellite image, where



(a) >32 km, 10.15-Hour M2P2 Data Collection across Different Locations: Maps show diverse environments such as lakeside trails, parks, and dense forests, highlighting the variety terrain and conditions in the dataset.



(b) LIO-SAM Mapping Results on Lake Braddock Trail: LiDAR point cloud (colored points) overlaid on a satellite image. Inset: LIO-SAM trajectory (blue) vs. raw GPS (green).

Fig. 7: Overview of the M2P2 dataset locations and mapping results. (a) shows various collection sites and terrain types. (b) shows LIO-SAM mapping results with comparison to raw GPS.

TABLE II: M2P2 Statistics

Attribute	Quantity
Total Size/Distance	≈2 TB / >32 km
Total Time/GPS Lock	10.15h / 8.93h
Average Speed	0.95 m/s
Number of RGB/Thermal Images	730606 / 361685
Number of Events	1.15×10^{11}
Number of Point Clouds	365297

point clouds align well with features like trail edges and vegetation. The inset shows the estimated trajectory (blue) compared to raw GPS (green); the latter suffers under tree cover, while LIO-SAM remains consistently accurate.

To support accurate sensor placement replication, we provide URDFs for the sensor suite on the Husky platform, along with calibrated transforms. Table II summarizes key statistics of the M2P2 dataset. Our synchronization scheme yields near-perfect alignment between RGB images and LiDAR point clouds, with only six mis-synchronized instances. The slightly lower count of thermal images is due to the camera's automatic shutter calibration, which pauses the stream for 0.4 seconds to correct non-uniformities.

VI. Experiment Results

We conduct three experiments using our M2P2 dataset to demonstrate its usefulness in off-road navigation and perception under degraded lighting conditions.

A. End-to-End Navigation Learning

To demonstrate the dataset's utility for end-to-end learning, we train a behavior cloning (BC) model to predict linear and angular velocities [35] from thermal images using ResNet-18. Since absolute temperature varies, we normalize each image using its min and max pixel values to obtain relative temperature. We deploy the BC model on a Husky robot for a 3.6 km autonomous run on a paved trail (Fig. 8). Lighting conditions vary from 235 lx to 0 lx (shown by path color), with most of the run in total darkness. The robot completes the route with only 11 human interventions, mostly due to thermal confusion between pavement and gravel. More robust navigation may require integrating other sensors, such as the event camera.

Fig. 8: Autonomous Navigation around a 3.6 km Trail with a BC model and Thermal Input: Lighting conditions drops from 255 lx at the beginning (light gray on the path, lower right) to 0 lx (black, upper left). 11 interventions (red crosses) are necessary to correct the robot when going off-course.

B. Perception in Degraded Visual Environments

To assess M2P2's utility for scene perception in degraded visual conditions, we compare metric depth estimation models. We train a 31M-parameter U-Net [36] to map thermal images to depth from LiDAR point clouds. Its performance is compared with DepthAnythingV2-Large [37], a monocular depth model with 335.3M parameters. As shown in Table III, U-Net achieves notably better results despite its smaller size, while DepthAnythingV2-Large struggles to generalize to the thermal domain.

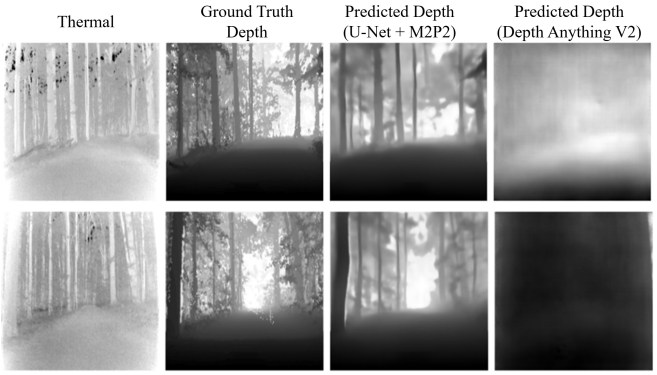


Fig. 9: Qualitative Depth Prediction Comparison on Unseen Data.

TABLE III: Quantitative Depth Prediction Comparison on Unseen Data.

Model	#Params (M) ↓	Abs Rel ↓	RMSE ↓	δ_1 ↑
DepthAnythingV2	335.3	0.66	8.43	0.03
U-Net + M2P2	31	0.13	2.12	0.82

Qualitative results in Fig. 9 support these findings. U-Net trained on M2P2 produces significantly higher-fidelity depth maps than DepthAnythingV2-Large. This underscores the value of domain-specific datasets like M2P2 for developing robust perception models in degraded visual conditions, where RGB-based methods often fail. Our results highlight the importance of such datasets in bridging the gap between standard perception and the challenges posed by non-traditional sensory inputs.

C. Passive Visual Odometry with Thermal and Event Data

A distinctive feature of M2P2 is the inclusion of calibrated, synchronized thermal and event data, enabling passive perception in extreme low-light. While prior work has explored visual-inertial odometry with RGB and event cameras [38], thermal-event fusion for odometry remains underexplored. This combination is promising for scenarios lacking visible light, such as nighttime off-road navigation or covert missions. RAMP-VO [38] is the closest prior work, and M2P2 contributes to advancing this research direction.

To showcase the potential of multi-modal fusion, we adapt the RAMP-VO framework originally for RGB and event data to work with thermal and event inputs from M2P2. We evaluate on a 157.5 m segment of the Burke Lake trail to test robustness under varying lighting. To simulate reduced light, we subsample the event stream, retaining 100%, 80%, 50%, and 25% of events, mimicking increasingly darker conditions. This setup allows us to analyze how thermal-event odometry performance degrades as event data becomes sparse.

Table IV presents the translational Absolute Trajectory Error (ATE) for each event subsampling level. As expected, the error generally increases as the event data becomes sparser.

TABLE IV: Translational ATE with Thermal-Event Fusion

Event Percentage	100%	80%	50%	25%
Translational ATE (m) ↓	8.79	11.60	12.79	12.49

VII. Conclusions and Future Work

This paper introduces M2P2, a novel multi-modal passive perception dataset specifically designed to address the challenges of off-road robot mobility in extreme low-light conditions. Unlike existing datasets, M2P2 uniquely combines thermal, event, and stereo RGB cameras, along with IMUs, GPS, and LiDAR for ground truth, providing a comprehensive representation of challenging off-road, low-light environments. We make the M2P2 dataset, along with our sensor suite design, publicly available to facilitate further research. We also present a robust multi-sensor calibration procedure, ensuring accurate data alignment across all modalities. Our initial experiments demonstrate that, even in complete darkness, off-road navigation, scene understanding, and vehicle state estimation are achievable using purely passive sensing.

While these initial experiments showcase the promise of individual modalities and limited fusion, the full realization of M2P2's potential requires deeper exploration of advanced sensor fusion techniques and their application to a wider range of mobility tasks. As the first step toward fully passive perception for off-road mobility in extreme low-light conditions, this work opens up a new avenue of future research. Some of the areas that could benefit from M2P2 include Visual Inertial Odometry [39]–[41], SLAM [42]–[44], and off-road kinodynamics modeling [45]–[49], all with the purely passive modalities available from our multi-modal sensor suite and dataset.

References

[1] X. Xiao, B. Liu, G. Warnell, and P. Stone, "Motion planning and control for mobile robot navigation using machine learning: a survey," *Autonomous Robots*, 2022.
[2] J. Hooks, M. S. Ahn, J. Yu, X. Zhang, T. Zhu, H. Chae, and D. Hong, "Alphred: A multi-modal operations quadruped robot for package delivery applications," *IEEE Robotics and Automation Letters*, 2020.
[3] L. Van Nguyen, S. Gibb, H. X. Pham, and H. M. La, "A mobile robot for automated civil infrastructure inspection and evaluation," in *2018 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*. IEEE, 2018.
[4] L. F. Oliveira, A. P. Moreira, and M. F. Silva, "Advances in agriculture robotics: A state-of-the-art review and challenges ahead," *Robotics*, 2021.
[5] R. R. Murphy, *Disaster robotics*. MIT press, 2014.
[6] U. Wandinger, "Introduction to lidar," in *Lidar: range-resolved optical remote sensing of the atmosphere*. Springer, 2005.
[7] L. Li *et al.*, "Time-of-flight camera—an introduction," *Technical white paper*, no. SLOA190B, 2014.
[8] A. J. Lee, Y. Cho, Y.-s. Shin, A. Kim, and H. Myung, "Vivid++: Vision for visibility dataset," *IEEE Robotics and Automation Letters*, 2022.
[9] J. Kim, H. Kim, S. Jeong, Y. Shin, and Y. Cho, "Diter++: Diverse terrain and multi-modal dataset for multi-robot slam in multi-session environments," *arXiv preprint arXiv:2412.05839*, 2024.
[10] M. Sivaprakasam, P. Maheshwari, M. G. Castro, S. Triest, M. Nye, S. Willits, A. Saba, W. Wang, and S. Scherer, "Tartandrive 2.0: More modalities and better infrastructure to further self-supervised learning research in off-road driving tasks," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024.

[11] K. Ebadi, Y. Chang, M. Palieri, A. Stephens, A. Hatteland, E. Heiden, A. Thakur, N. Funabiki, B. Morrell, S. Wood *et al.*, "LAMP: Large-scale autonomous mapping and positioning for exploration of perceptually-degraded subterranean environments," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020.

[12] R. Thakker, N. Alatur, D. D. Fan, J. Tordesillas, M. Paton, K. Otsu, O. Toupet, and A.-a. Agha-mohammadi, "Autonomous off-road navigation over extreme terrains with perceptually-challenging conditions," in *Experimental Robotics: The 17th International Symposium*. Springer, 2021.

[13] Y. Chang, K. Ebadi, C. E. Denniston, M. F. Ginting, A. Rosinol, A. Reinke, M. Palieri, J. Shi, A. Chatterjee, B. Morrell, A.-a. Aghamohammadi, and L. Carlone, "LAMP 2.0: A robust multi-robot slam system for operation in challenging large-scale underground environments," *IEEE Robotics and Automation Letters*, 2022.

[14] M. Wermelinger, P. Fankhauser, R. Diethelm, P. Krüsi, R. Siegwart, and M. Hutter, "Navigation planning for legged robots in challenging terrain," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2016.

[15] L. Sharma, M. Everett, D. Lee, X. Cai, P. Osteen, and J. P. How, "RAMP: A risk-aware mapping and planning pipeline for fast off-road ground robot navigation," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023.

[16] C. Chung, G. Georgakis, P. Spieler, C. Padgett, A. Agha, and S. Khattak, "Pixel to elevation: Learning to predict elevation maps at long range using images for autonomous offroad navigation," *IEEE Robotics and Automation Letters*, 2024.

[17] L. Wellhausen, A. Dosovitskiy, R. Ranftl, K. Walas, C. Cadena, and M. Hutter, "Where should i walk? predicting terrain properties from images via self-supervised learning," *IEEE Robotics and Automation Letters*, 2019.

[18] P. Fankhauser, M. Bloesch, and M. Hutter, "Probabilistic terrain mapping for mobile robots with uncertain localization," *IEEE Robotics and Automation Letters*, 2018.

[19] M. Wigness, S. Eum, J. G. Rogers, D. Han, and H. Kwon, "A rugd dataset for autonomous navigation and visual perception in unstructured outdoor environments," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019.

[20] X. Meng, N. Hatch, A. Lambert, A. Li, N. Wagener, M. Schmittle, J. Lee, W. Yuan, Z. Chen, S. Deng *et al.*, "Terrainnet: Visual modeling of complex terrain for high-speed, off-road navigation," *arXiv preprint arXiv:2303.15771*, 2023.

[21] P. Sermanet, R. Hadsell, M. Scoffier, U. Muller, and Y. LeCun, "Mapping and planning under uncertainty in mobile robots with long-range perception," in *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2008.

[22] M. Bajracharya, J. Ma, M. Malchano, A. Perkins, A. A. Rizzi, and L. Matthies, "High fidelity day/night stereo mapping with vegetation and negative obstacle detection for vision-in-the-loop walking," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2013.

[23] M. G. Castro, S. Triest, W. Wang, J. M. Gregory, F. Sanchez, J. G. Rogers, and S. Scherer, "How does it feel? self-supervised costmap learning for off-road vehicle traversability," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023.

[24] X. Cai, M. Everett, J. Fink, and J. P. How, "Risk-aware off-road navigation via a learned speed distribution map," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022.

[25] G. Kahn, P. Abbeel, and S. Levine, "BADGR: An autonomous self-supervised learning-based navigation system," *IEEE Robotics and Automation Letters*, 2021.

[26] S. Jeong, H. Kim, and Y. Cho, "Diter: Diverse terrain and multi-modal dataset for field robot navigation in outdoor environments," *IEEE Sensors Letters*, pp. 1–4, 03 2024.

[27] P. Jiang, P. Osteen, M. Wigness, and S. Saripalli, "Rellis-3d dataset: Data, benchmarks and analysis," in *2021 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2021.

[28] A. Z. Zhu, D. Thakur, T. Özaslan, B. Pfrommer, V. Kumar, and K. Daniilidis, "The multivehicle stereo event camera dataset: An event camera dataset for 3d perception," *IEEE Robotics and Automation Letters*, 2018.

[29] J. Delmerico, T. Cieslewski, H. Rebecq, M. Faessler, and D. Scaramuzza, "Are we ready for autonomous drone racing? the uzh-fpv drone racing dataset," in *2019 International Conference on Robotics and Automation (ICRA)*, 2019.

[30] W. Maddern and S. Vidas, "Towards robust night and day place recognition using visible and thermal imaging," in *Proceedings of the RSS 2012 Workshop: Beyond laser and vision: Alternative sensing techniques for robotic perception*. University of Sydney, 2012.

[31] Y. Choi, N. Kim, S. Hwang, K. Park, J. S. Yoon, K. An, and I. S. Kweon, "Kaist multi-spectral day/night data set for autonomous and assisted driving," *IEEE Transactions on Intelligent Transportation Systems*, 2018.

[32] K. Chaney, F. Cladera, Z. Wang, A. Bisulco, M. A. Hsieh, C. Korpela, V. Kumar, C. J. Taylor, and K. Daniilidis, "M3ed: Multi-robot, multi-sensor, multi-environment event dataset," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2023.

[33] M. Muglikar, M. Gehrig, D. Gehrig, and D. Scaramuzza, "How to calibrate your event camera," in *IEEE Conf. Comput. Vis. Pattern Recog. Workshops (CVPRW)*, June 2021.

[34] P. Furgale, J. Rehder, and R. Siegwart, "Unified temporal and spatial calibration for multi-sensor systems," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2013.

[35] A. Datar, C. Pan, M. Nazeri, and X. Xiao, "Toward wheeled mobility on vertically challenging terrain: Platforms, datasets, and algorithms," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024.

[36] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*. Springer, 2015.

[37] L. Yang, B. Kang, Z. Huang, Z. Zhao, X. Xu, J. Feng, and H. Zhao, "Depth anything v2," *arXiv preprint arXiv:2406.09414*, 2024.

[38] R. Pellerito, M. Cannici, D. Gehrig, J. Belhadj, O. Dubois-Matra, M. Casasco, and D. Scaramuzza, "Deep visual odometry with events and frames," in *IEEE/RSJ International Conference on Intelligent Robots (IROS)*, June 2024.

[39] Z. Zhang and D. Scaramuzza, "A tutorial on quantitative trajectory evaluation for visual (-inertial) odometry," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018.

[40] Z. Huai and G. Huang, "Robocentric visual–inertial odometry," *The International Journal of Robotics Research*, 2022.

[41] M. Bloesch, S. Omari, M. Hutter, and R. Siegwart, "Robust visual inertial odometry using a direct ekf-based approach," in *2015 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE, 2015.

[42] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, "Monoslam: Real-time single camera slam," *IEEE transactions on pattern analysis and machine intelligence*, 2007.

[43] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "Orb-slam: a versatile and accurate monocular slam system," *IEEE transactions on robotics*, 2015.

[44] T. Taketomi, H. Uchiyama, and S. Ikeda, "Visual slam algorithms: A survey from 2010 to 2016," *IPSJ transactions on computer vision and applications*, 2017.

[45] H. Karnan, K. S. Sikand, P. Atreya, S. Rabiee, X. Xiao, G. Warnell, P. Stone, and J. Biswas, "Vi-ikd: High-speed accurate off-road navigation using learned visual-inertial inverse kinodynamics," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022.

[46] P. Atreya, H. Karnan, K. S. Sikand, X. Xiao, S. Rabiee, and J. Biswas, "High-speed accurate robot control using learned forward kinodynamics and non-linear least squares optimization," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022.

[47] A. Datar, C. Pan, and X. Xiao, "Learning to model and plan for wheeled mobility on vertically challenging terrain," *arXiv preprint arXiv:2306.11611*, 2023.

[48] A. Datar, C. Pan, M. Nazeri, A. Pokhrel, and X. Xiao, "Terrain-attentive learning for efficient 6-dof kinodynamic modeling on vertically challenging terrain," in *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2024.

[49] A. Pokhrel, A. Datar, M. Nazeri, and X. Xiao, "CAHSOR: Competence-aware high-speed off-road ground navigation in SE (3)," *IEEE Robotics and Automation Letters*, 2024.