ViViD : Vision for Visibility Dataset

Alex Junho Lee[1], Younggun Cho[1], Sungho Yoon[2], Youngsik Shin[1], and Ayoung Kim[1,2*]

***Abstract*— In this paper, we provide a dataset capturing unconventional visual data obtained from poor lighting conditions. Our data provides normal and poor illumination sequences recorded by thermal, depth, and temporal difference sensor for indoor and outdoor trajectories. For a visual-inertial navigation system free from lighting conditions, we suggest obtaining visual information from a domain out of comparing visible light intensity. We present a dataset collected from co-aligned alternative cameras with inertial measurements. Our sensor set collects data independently from visible light intensity by measuring the amount of infrared dissipation, depth by structured reflection and instantaneous temporal changes in luminance. We provide these measurements along with inertial sensors and ground truth, for visual-inertial navigation tasks under poor illumination.**

I. Introduction

With recent interests in autonomous navigation, simultaneous localization and mapping (SLAM) is becoming an important topic for its fundamental role in localization and object recognition. Visual algorithms using images to solve SLAM problems, have been a large part from its sensor availability and intuitiveness. In developing robust long-term navigation technique, it is crucial to overcome lighting and motion disturbances from environmental conditions. Variances due weather changes on the road, large luminance modification in indoor, and even in disastrous circumstances are the key factors robots are required to overcome. However, from the limitation of scene illustration hindered by the frame rate and low dynamic range in cameras, it had been challenging to solve SLAM problems in natural conditions with typical cameras. Thus, a large variety of test environments was necessary to develop robust visual algorithms in the real world, and it was solved by public data.

Public datasets provide large variations of environment and sensor characteristics. Some of datasets contain indoor (TUM RGB-D [1], EuRoC [2]) or outdoor scenes (KITTI [3], Cityscapes [4], Complex Urban (KAIST) [5]), including challenging environments such as low light or time-varying luminance conditions. Synthetic datasets (ICL-NUIM [6], SceneNet RGB-D [7], InteriorNet [8]) are also available, providing more ease in solving spatial perception and mapping problems in noisy environments. However, it is still challenging to achieve robust visual SLAM in every low-visibility situation. For typical cameras collecting information by integrating photons in the visible spectrum for a fixed exposure time, the captured image is highly dependent on the external light source. Thus classical cameras tend to concentrate more on lighting than the texture itself, making the camera vulnerable in many natural circumstances. From ideas of collecting information from other domains than light intensity, the multi-spectral or dynamic vision systems were introduced. These type of sensors collects information from infrared radiation or encodes log differences in luminance, making themselves more independent from external lighting conditions. However, since the outputs of these sensors are different from typical cameras, they requires modifications on classical visual algorithms to robustly solve SLAM problem in various environments.

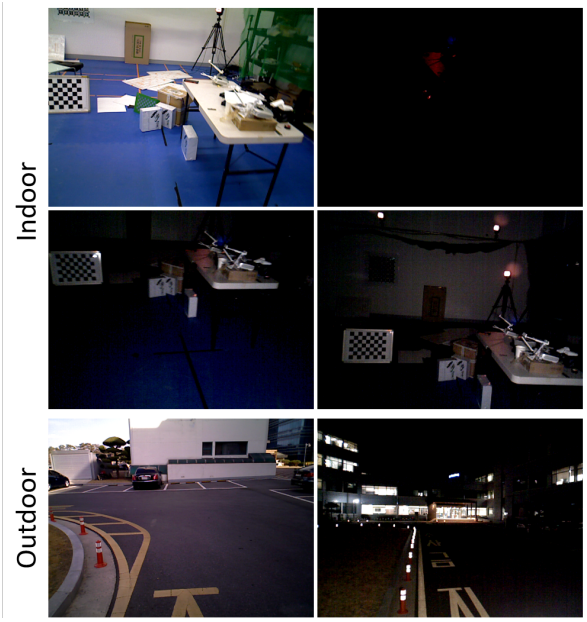


Fig. 1. Sample RGB images. The dataset was prepared in varying light condition targetting dynamic and irregular ego-motion under a low light condition.

Therefore, in this paper, we provide alternative vision sensor measurements for developing robust visual SLAM independent of external light changes. Our dataset consists of RGB, depth, thermal, and event measurements on indoor and outdoor sequences, including normal to extremely low lighting and even time-varying light conditions.

II. Related Works

Numbers of datasets for benchmarking SLAM has been newly introduced [4], [5], [9] with clear sight and high visibility. Meanwhile, in the real world, lighting conditions are often uncooperative, making computer vision algorithms fail.

[1]Department of Civil and Environmental Engineering, KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea [alex_jhlee,yg.cho,youngsik,ayoungk]@kaist.ac.kr

[2]Robotics Program, KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea sungho.yoon@kaist.ac.kr

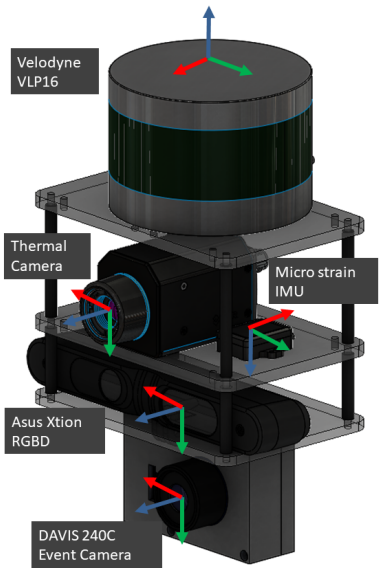


Fig. 2. Sensors hardware configuration.

Moreover, there are few of datasets made to test experimental environments with environmental variations.

NCLT [10] and TUM MonoVO [11] introduced large scale data with huge variance in environments for long-term visual SLAM. These datasets cover challenging indoor and outdoor sequences including natural light and weather changes. However, obtaining the limited information from the classical camera restricts the bandwidth of data from the environment, skipping potentially important information.

By using other types of visual sensors, Choi et al. [12] presented a multi-spectral day/night dataset with the sensor set of stereo RGB, LiDAR and a thermal camera. Their data contains variance along the whole day. Furthermore, numbers of labeled data are provided for autonomous navigation. However, camera movements in the dataset are limited to planar motion because the system is mounted on a car.

In [13], the authors have presented a dataset measuring various lighting and motion sequences with two event cameras aligned with inertial sensors and an inertial measurement unit (IMU). The dataset contains a large variety of condition changes, suggesting utilizing the event camera for low latency and high dynamic range characteristics. However, due to the low resolution of event cameras compared to the thermal or RGB-D camera, the resulting estimation runs the risk of not fully achieving desirable accuracy.

As listed above, several datasets have contained environmental variations with different sensor sequences. However, none of the datasets have full potential for covering any motion in an unspecified natural environment. In this paper, we propose a dataset to record both thermal and events with depth in order to deal with two significant disturbances: luminance conditions and motion.

III. Datasets

A. Sensors and Data types

To capture visual information even under insufficient lighting condition, three cameras were installed along with the inertial measurement unit as in Figure 2. Through this sensor system, we aim to measure visual information free from external light conditions. These unique sensors collect each data from infrared radiation, which is dependent on the temperature of the object, structured light depth measurement, and relative intensity changes not absolute.

The dataset is provided in binary format in rosbag, while the topic lists with specifications can be found in Table I. Note that in the thermal camera, the format of the image is not a typical 8-bit int but 14 bits enclosed in 16 bits.

B. Sequence Description

The full data list is detailed in Table II. The sequences are composed of two distinct locations (indoor / outdoor) and four different light conditions (normal / dark / dimmed / varying light) with motion variances for indoor sequences. The reconstructed structure and images of each set can be found in Figure 4 and Figure 5.

1) Environments: The first and second batches of the dataset are recorded in different locations. In indoor sequences, global poses of the platform are captured via motion capture system. The scale of the room is $12.3\ \mathrm{m} \times 8.9\ \mathrm{m} \times 4.5\ \mathrm{m}$ with 12 cameras mounted on the wall for motion capture. The system uses infrared strobes to track the reflected markers of desired platforms.

For outdoor sequences, a pose obtained with LeGO-LOAM [14] was used as a ground truth. The location of the outdoor trajectory is nearly enclosed by buildings, lowering the accuracy of Global Positioning System (GPS) rather appropriate for LiDAR-based algorithms. The total size of the enclosed region is about $60\ \mathrm{m} \times 40\ \mathrm{m}$, and the trajectory length is around $50\ \mathrm{m}$.

2) Illumination and Ego-motion Variance: In each batch, light and motion variance was applied to the enclosure's natural disturbance circumstances : in the real world, robots experience both uncooperative luminance and abrupt ego-motion. The dataset consists of three sequences from normal, dark and controlled illumination conditions. For indoor environments, rapid movements were recorded assuming drone tracking or hand-held scenarios.

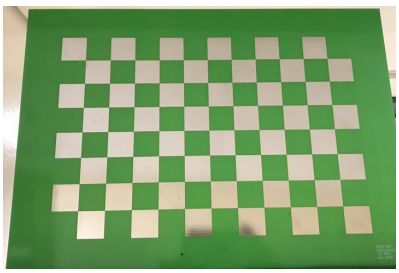


Fig. 3. Checkerboard pattern printed with metal on circuit board.

TABLE I

SENSOR SPECIFICATIONS AND DATA TYPES

Sensors	Specifications	Topic name	Description	Message type
Thermal	FLIR A65 640×512 pixel, 20Hz FOV : 90°vert., 69°horiz. Spectral Range : 7.5-13 μm	/thermal/camera_info /thermal/image_raw	Header Image (14bit, 1ch)	sensor_msgs/CameraInfo 640×512 uint16 Image
RGB-D	Asus Xtion Pro Live 640×480 pixel, 30Hz FOV : 45°vert., 58°horiz.	/camera/depth/camera_info /camera/depth/camera_image /camera/rgb/camera_info /camera/rgb/camera_image	Header Image (8bit, 1ch) Header Image (8bit, 3ch)	sensor_msgs/CameraInfo 640×480 uint8 Image sensor_msgs/CameraInfo 640×480 uint8 Image
Event	DAVIS 240C 240×180 pixel, upto 12 MEPS FOV : 40.4°vert., 52.3°horiz. IMU : MPU 6150	/dvs/events	Event	dvs_msgs/EventArray
Inertial	On-board IMU (MPU 6150) Microstrain IMU (3DM-GX5-25)	/dvs/imu /imu/data	Imu	sensor_msgs/Imu
LiDAR	Velodyne VLP-16	/velodyne_pointcloud	Pointcloud	sensor_msgs/PointCloud2

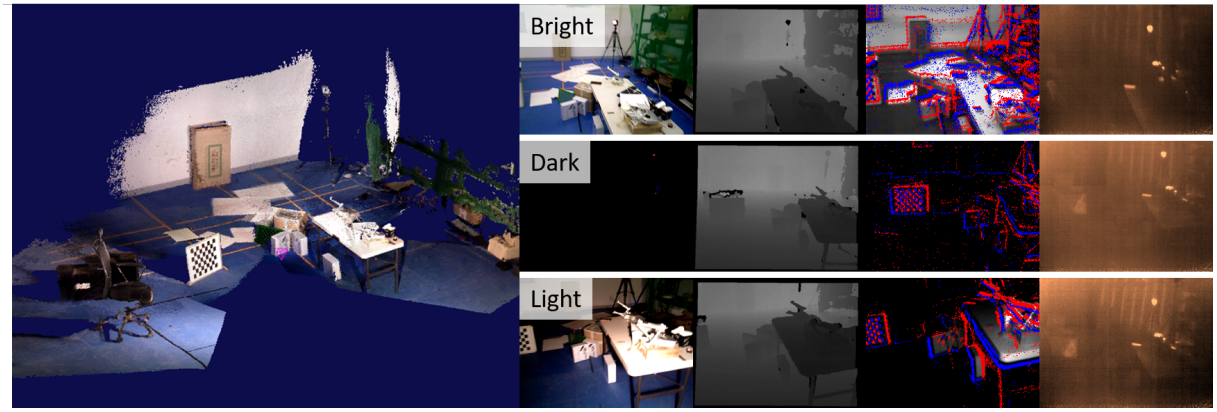


Fig. 4. Right shows the reconstructed 3D map. Left shows indoor samples with garnered under bright, dark and controlled illumination.

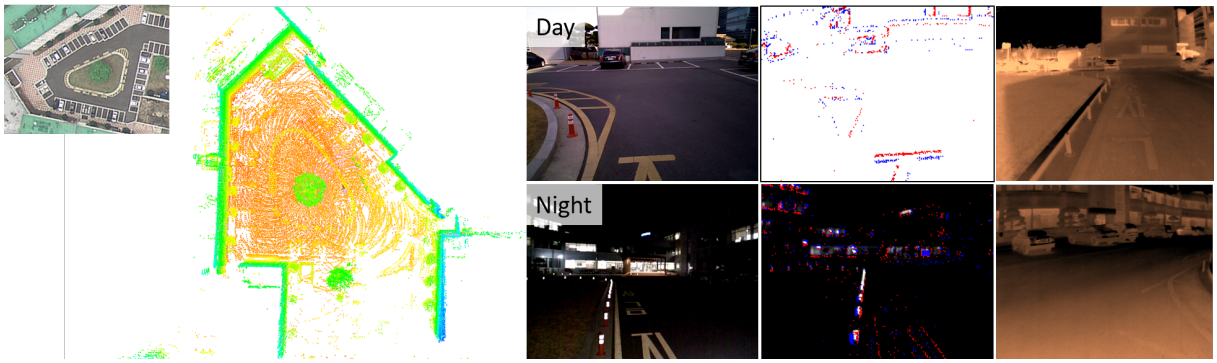


Fig. 5. Right shows the reconstructed 3D map. Left shows indoor samples from day and night.

IV. CALIBRATION AND SYNCHRONIZATION

A. Calibration

For calibration, we offer a multimodal relative pose of the base point to RGB-D, RGB-D to IMU, RGB-D to thermal, and RGB-D to an event camera. We calibrated the relative poses with the Kalibr toolbox [15], [16], [17] and checkerboard pattern made of aluminum on printed circuit board as in Figure 3. Since the thermal camera only detects infrared radiation from objects, we dissipated heat in advance to the aluminum pattern to get proper contrast. Further processes are then done with the toolbox; we provide focal length, principal point of each camera with distortion parameters from the calibrated results.

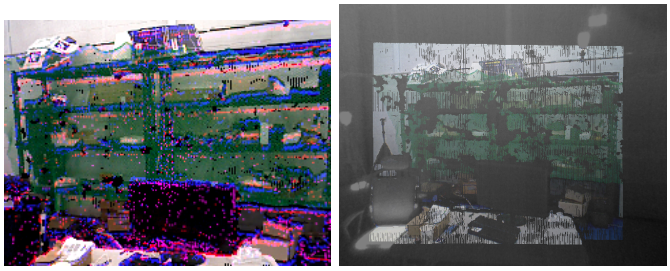
The extrinsic between the base point and cameras was done with CamOdoCal [18]. Since we measured the pose of the tracking points, not the sensor base, we needed to obtain the relative poses of the camera. By exploiting the motion capture system, we moved our sensors at known positions of AprilTags [19] and calculated the cameras' relative poses.

For timestamps in the event camera, we calculated the temporal offset between the global pose and local frames by

TABLE II

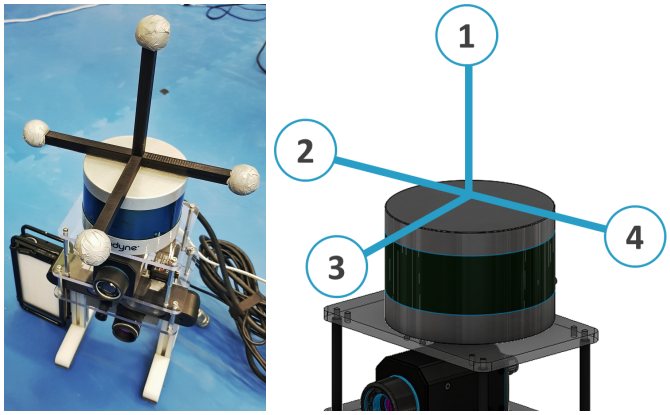
ENVIRONMENT SETTING FOR EACH SEQUENCES

Sequence	Ambient Light	Additional Light	Motion	Pose GT
Indoor	Bright	OFF	Robust	Vicon
	Bright	OFF	Fast	Vicon
	Dark	OFF	Robust	Vicon
	Dark	OFF	Fast	Vicon
	Dark	ON	Robust	Vicon
	Dark	ON	Fast	Vicon
Outdoor	Bright	OFF	Robust	LOAM
	Dark	OFF	Robust	LOAM
	Dark	ON	Robust	LOAM



(a) RGB-D pointcloud projected into accumulated events for fixed time. (b) RGB-D pointcloud projected into thermal image.

Fig. 6. Overlapped images from results of extrinsic calibration. Each pointcloud is transformed into the other camera's view, using depth information from RGB-D camera.



(a) Sensor jig with GT markers (b) GT marker coordinates

Fig. 7. Sensor configurations for ground-truth and references. We utilized VICON and LiDAR for reference pose generations.

optimizing the correlation among angular velocity measurements from the global pose and event camera's IMU. We assumed a fixed time offset for each sequence and modified event timestamps to compensate. For overlooking calibration results, overlapped images of RGB-D and other cameras were presented in Figure 6.

B. Ground Truth Generation

We used the Cortex motion capture system called KARPE (KAIST Arena w/ Real-time Positioning Environment) [20] to generate the ground-truth and attached the markers to the sensor jig as seen in Figure 7(a) (body frame). The motion of the sensor jig's body frame from the world frame of the motion capture system was recorded at 100 Hz. Since we know the cam2body calibration, we can get the ground-truth relative pose of each sensor.

Since the motion capture system can not be used in the outdoor environment, LeGO-LOAM [14], which is the latest LiDAR SLAM method, is used as the baseline. To obtain the LiDAR measurement, a VLP16 LiDAR was attached to the upper layer of the sensor jig. Following the previous work [21], the extrinsic calibration of LiDAR to the camera is obtained by plane matching in RGB-D depth and LiDAR frame.

V. CONCLUSION

We present a dataset to overcome poor lighting conditions, by providing 6-degree of freedom (DOF) ground truth poses and measurements above the visible light spectrum by depth, radiation image, and events. We hope our dataset will enable the comparison of robust algorithms for robot navigation problems, regardless of motion or environments.

REFERENCES

[1] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, "A benchmark for the evaluation of rgb-d slam systems," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Oct. 2012.

[2] M. Burri, J. Nikolic, P. Gohl, T. Schneider, J. Rehder, S. Omari, M. W. Achtelik, and R. Siegwart, "The euroc micro aerial vehicle datasets," *International Journal of Robotics Research*, 2016.

[3] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the KITTI vision benchmark suite," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2012.

[4] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 3213–3223.

[5] J. Jeong, Y. Cho, Y.-S. Shin, H. Roh, and A. Kim, "Complex urban dataset with multi-level sensors from highly diverse urban environments," in *The International Journal of Robotics Research*, 2019.

[6] A. Handa, T. Whelan, J. McDonald, and A. Davison, "A benchmark for RGB-D visual odometry, 3D reconstruction and SLAM," in *IEEE Intl. Conf. on Robotics and Automation, ICRA*, Hong Kong, China, May 2014.

[7] J. McCormac, A. Handa, S. Leutenegger, and A. J. Davison, "Scenenet rgb-d: Can 5m synthetic images beat generic imagenet pre-training on indoor segmentation?" in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2678–2687.

[8] W. Li, S. Saeedi, J. McCormac, R. Clark, D. Tzoumanikas, Q. Ye, Y. Huang, R. Tang, and S. Leutenegger, "Interiornet: Mega-scale multi-sensor

photo-realistic indoor scenes dataset," *arXiv preprint arXiv:1809.00716*, 2018.

[9] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The kitti dataset," *International Journal of Robotics Research*, 2013.

[10] N. Carlevaris-Bianco, A. K. Ushani, and R. M. Eustice, "University of Michigan North Campus long-term vision and lidar dataset," *International Journal of Robotics Research*, vol. 35, no. 9, pp. 1023–1035, 2015.

[11] J. Engel, V. Usenko, and D. Cremers, "A photometrically calibrated benchmark for monocular visual odometry," *arXiv preprint arXiv:1607.02555*, 2016.

[12] 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*, vol. 19, no. 3, pp. 934–948, 2018.

[13] 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*, vol. 3, no. 3, pp. 2032–2039, 2018.

[14] 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.

[15] 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, pp. 1280–1286.

[16] P. Furgale, T. D. Barfoot, and G. Sibley, "Continuous-time batch estimation using temporal basis functions," in *2012 IEEE International Conference on Robotics and Automation*. IEEE, 2012, pp. 2088–2095.

[17] J. Maye, P. Furgale, and R. Siegwart, "Self-supervised calibration for robotic systems," in *2013 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2013, pp. 473–480.

[18] L. Heng, B. Li, and M. Pollefeys, "Camodocal: Automatic intrinsic and extrinsic calibration of a rig with multiple generic cameras and odometry," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2013, pp. 1793–1800.

[19] E. Olson, "Apriltag: A robust and flexible visual fiducial system," in *2011 IEEE International Conference on Robotics and Automation*, May 2011, pp. 3400–3407.

[20] H.-Y. Kim, J.-S. Lee, H.-L. Choi, and J.-H. Han, "Autonomous formation flight of multiple flapping-wing flying vehicles using motion capture system," *Aerospace Science and Technology*, vol. 39, pp. 596–604, 2014.

[21] Y. Bok, Y. Jeong, D.-G. Choi, and I. S. Kweon, "Capturing village-level heritages with a hand-held camera-laser fusion sensor," *International Journal of Computer Vision*, vol. 94, no. 1, pp. 36–53, Aug 2011.