

# STheReO: Stereo Thermal Dataset for Research in Odometry and Mapping

Seungsang Yun<sup>1</sup>, Minwoo Jung<sup>1</sup>, Jeongyun Kim<sup>1</sup>, Sangwoo Jung<sup>1</sup>,  
Younghun Cho<sup>2</sup>, Myung-Hwan Jeon<sup>3</sup>, Giseop Kim<sup>4</sup> and Ayoung Kim<sup>1\*</sup>



Fig. 1: The stereo thermal image pairs (left) and the thermal 3D point cloud maps (right) for three different times for our KAIST sequences are visualized. The 3D point cloud map at the middle is constructed using our SE(3) baseline trajectory. Our STheReO dataset provides stereo thermal camera images with multiple commonly used sensors such as IMU, RGB camera, and LiDAR. Using our bundle of sensors and baseline trajectories, we encourage exploring stereo thermal camera-based SLAM researches that can be used for robotic applications such as thermal-LiDAR mapping as in the right plots and Fig. 7. From the thermal LiDAR map (right), we can see the 3D temperature variation within a day when a particular building (red box) in the morning heats up faster than others as it faces the direction of the sunrise. We can also see the distinct thermal intensity gaps among different times for the same site.

**Abstract**—This paper introduces a stereo thermal camera dataset (STheReO) with multiple navigation sensors to encourage thermal SLAM researches. A thermal camera measures infrared rays beyond the visible spectrum therefore it could provide a simple yet robust solution to visually degraded environments where existing visual sensor-based SLAM would fail. Existing thermal camera datasets mostly focused on monocular configuration using the thermal camera with RGB cameras in a visually challenging environment. A few stereo thermal rig were examined but in computer vision perspective without supporting sequential images for state estimation algorithms. To encourage the academia for the evolving stereo thermal SLAM, we obtain nine sequences in total across three spatial locations and three different times per location (e.g., morning,

day, and night) to capture the variety of thermal characteristics. By using the STheReO dataset, we hope diverse types of researches will be made, including but not limited to odometry, mapping, and SLAM (e.g., thermal-LiDAR mapping or long-term thermal localization). Our datasets are available at <https://sites.google.com/view/rpmsthereo/>.

## I. INTRODUCTION

State estimation of robot poses and the surrounding landmark locations (well known as simultaneous localization and mapping (SLAM) [8]) is an essential task for a mobile robot to navigate an unknown environment. Particularly, for some harsh environments where visual information is easily degraded (e.g., subterranean [5] or nighttime [9]), the robust perception should be guaranteed for safe navigation. Recently, thermal-infrared cameras [10, 11, 12] have been employed to cope with such visually degraded environments where conventional gray or RGB cameras cannot provide rich information thus the existing visual SLAM algorithms [13, 14, 15, 16] may fail.

However, thermal camera-based SLAM algorithms [9, 17] have been overlooked compared to conventional cameras and

<sup>1</sup>S. Yun, M. Jung, J. Kim, S. Jung, and A. Kim are with the Dept. of Mech. Eng., SNU, S. Korea [seungsang, moonshot, jeongyunkim, dan0130, ayoungk]@snu.ac.kr

<sup>2</sup>Y. Cho is with the Dept. of Civil and Env. Eng., KAIST, S. Korea lucascho@kaist.ac.kr

<sup>3</sup>M. Jeon is with the Robotics Program, KAIST, Daejeon, S. Korea, myunghwan.jeon@kaist.ac.kr

<sup>4</sup>G. Kim is affiliated with Autonomous Driving Group, NAVER LABS, S. Korea giseop.kim@naverlabs.com

\*This research was supported NRF (2020R1C1C1006620) and IITP (2022-0-00480) Korea.

TABLE I: Comparison with other datasets including thermal camera

Datasets	SLAM Research Relevancy	Thermal - Stereo	Thermal - Monocular	LiDAR	IMU	GPS	RGB
PST900 [1]			✓				✓
ComplexUrban [2]	✓			✓	✓	✓	✓
KITTI [3]	✓			✓	✓	✓	✓
ViViD [4]	✓		✓	✓	✓		✓
SubT [5]	✓		✓	✓	✓		✓
CATS [6]		✓	✓	✓	✓	✓	✓
KAIST Day/Night [7]	✓		✓	✓	✓	✓	✓
Ours (STheReO)	✓	✓	✓	✓	✓	✓	✓

LiDAR-based methods [18, 19, 20]. A few datasets are currently available that have thermal image streams designed for SLAM researches (i.e., providing various traversals for multiple environments with ground-truth trajectories)[7, 4, 5]. Unfortunately, none of them supports stereo thermal camera rigs for stereo thermal image-based SLAM researches. CATS [6] provides the stereo configuration but differs from us in the target domain without providing a sequential data for SLAM.

A thermal camera inevitably accompanies non-uniformity correction (NUC) [21] to mitigate sensor noise that internal heats accumulate. NUC freezes the camera and drops image frames for a few seconds, which could degrade state estimation algorithm performance (e.g., tracking lost). Potential alleviation is to force the NUC off during image acquisition as in [9] which may cause gradual temperature drift. With the NUC on, perfect synchronization among thermal cameras is hardly achievable. This asynchronous characteristic of the data would be the critical challenges [22] for most visual SLAM that assumes synchronized image stream.

Together with the stereo configuration, a dataset containing enough NUC events is required to develop a robust SLAM algorithm. To the best of our knowledge, there is no publicly released stereo thermal dataset relevant to SLAM. In this paper, we provide a dataset particularly designed for thermal camera-based researches in odometry and mapping in real-world outdoor environments under temporal variations.

- To the best of our knowledge, our dataset (STheReO) is the first one equipped with **stereo thermal cameras** with multiple sensors in SLAM academia, as seen in Fig. 1.
- We provide a set of multiple sequences that cover three different spatial sites under various temperature and illumination conditions (e.g., morning, day, and night), and the NUC occurrence diversities.
- We provide SE(3) baseline trajectories for all sequences for the state estimation algorithm evaluation.

By releasing this STheReO dataset, we hope to accelerate **Stereo Thermal Researches** in not only **Odometry** and **mapping** but also thermal place recognition or long-term autonomy.

## II. RELATED WORKS

### A. Thermal Camera for SLAM

Existing thermal camera-based state estimation methods fused the thermal data with common other sensor suites such as IMU [17], LiDAR [9], or stereo pair [23]. Mouats et al. [23] proposed stereo thermal camera-based odometry. Khattak et al. [17] proposed a direct visual-inertial odometry method by avoiding the transformation of 14-bit raw thermal data to an 8-bit image, which caused information loss but generally required for indirect methods. Shin and Kim [9] proposed sparse depth-enhanced thermal camera SLAM by exploiting a 3D LiDAR. Recently, deep learning-based thermal SLAM researches have been proposed [10, 11, 12]. For thermal camera in/extrinsic calibration, Saponaro et al. [24] used a chessboard pattern with heterogenous materials (to be selectively heated) to measure the temperature differences and capture the corner points in the chessboard. However, this method is limited in obtaining a sharp image due to heat transfer. By contrast, [25] used a calibration board containing a set of square holes to obtain more sharpened patterns from the image. Meanwhile, Borges and Vidas [26] remarked on the importance of NUC management for the practical infrared visual odometry.

Despite these existing studies, we believe the scale of thermal camera-based SLAM researches is still smaller than those of common visual SLAM. This scale limitation may be because the aforementioned methods were all evaluated on their own private datasets, making it difficult to be quantitatively compare them in equivalent settings.

### B. Datasets for Thermal SLAM research

Despite a few datasets including thermal cameras, currently accessible datasets mostly targeted for the front-end perception such as pedestrian detection [27] and segmentation [1], which is less suitable for SLAM research without providing temporally sequential data stream.

Moreover, publicly available SLAM-relevant datasets equipped with thermal cameras are less developed [7, 4, 5] compared to the common (gray or RGB) visual domain [3, 28, 29, 30, 31, 2, 32] or range sensors [33, 34]. Choi et al. [7] published a dataset including a monocular thermal camera, stereo RGB, 3D LiDAR, and IMU sensors. They used a beam splitter to acquire parallax-free pair data between

TABLE II: The sensor specification of our dataset

Sensor	Manufacturer, Model	Description	Hz
Thermal Camera	FLIR, A-65	0.9° and 0.06 m resolution, 90° Vert., 69° horiz. FOV, (640,512) resolution	10
RGB Camera	FLIR, Flea-3	Global shutter 3channel camera, (1280,560) resolution	10
3D LiDAR	Ouster, OS1-128	128 channel, 360° FOV, (1024,128) resolution	10
IMU	Xsens, MTi-300	High performing Attitude and Heading Reference System	100
RTK GPS	Novatel, CPT-7	Providing accurate pose using GNSS + INS solution	100

RGB and thermal images. However, they designed the dataset to focus more on perception (e.g., object detection, vision sensor enhancement, depth estimation, and multi-spectral colorization) than state estimation. Rogers et al. [5] provided a monocular thermal camera dataset with multi-sensors in the subterranean tunnel environment. This dataset does not provide the thermal camera’s intrinsic and extrinsic parameters with respect to other sensors. A monocular thermal camera and its calibration parameters are available within the ViViD dataset [4], however, the spatial and temporal scales of the dataset are restricted compared to ours. A more detailed comparisons with other datasets are arranged in Table I.

In this paper, we provide the first stereo thermal camera dataset with multi-sensors and 6D ground-truth trajectories. Our dataset consists of a stereo thermal camera and navigation sensors appropriate to SLAM researches. Our sequences include large-scale spatial coverages for three distinct locations at three different times (i.e., morning, day, and night). We also provide thermal cameras’ extrinsic parameters, which is calculated by ground extraction-based targetless calibration, to boost developing sensor-fusion SLAM with our stereo thermal cameras.

### III. SYSTEM OVERVIEW

#### A. Sensor Configuration

Our sensors and the hardware system are visualized in Fig. 2. The sensor-rig is placed on top of a car platform. We have multiple sensors commonly used for SLAM such as RGB cameras, LiDAR, IMU, and RTK-GPS as well as the stereo thermal cameras. Table II represents the sensor’s specifications. The baselines of the stereo thermal cameras and the stereo RGB cameras are 0.62 m and 0.50 m, respectively. We note the stereo thermal cameras are located outer than the RGB cameras to maximize their baseline. Each sensor’s coordinates are drawn as colored arrows. Please refer to the details of the parameters from Fig. 2 and our dataset site<sup>1</sup>.

#### B. Thermal Camera Intrinsic Calibration

Unlike RGB cameras that measure visible light, a thermal camera measures infrared rays related to the object’s surface temperature. Therefore, blurs in thermal image are inevitable due to heat transfer. To mitigate this issue, existing thermal camera intrinsic calibration methods [24] use the calibration board made of different materials to obtain heat differences. However, even in this method, heat transfer occurs and blurry images cannot be avoided. Thus, by following the work of

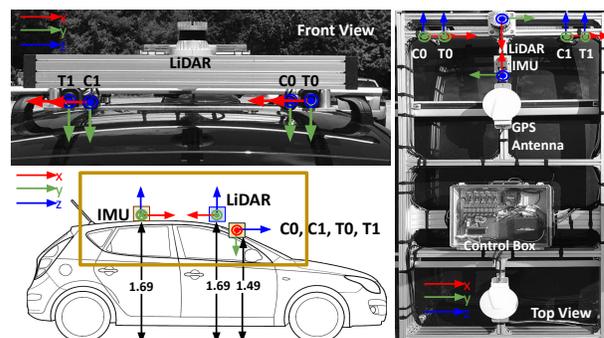


Fig. 2: The STheReO dataset’s sensor configuration. The upper left picture shows the sensor system’s front view, and both the cameras and LiDAR are mounted to face the front of the vehicle. The sensors’ x, y, and z coordinates are represented as red, blue, and green arrows, respectively. The lower left figure shows the sensors’ height, and the right figure shows the top view of the sensor system.

[25], we successfully found the exact corner points in thermal images and conduct the intrinsic parameter calibration using a pattern board with square-shaped holes. A computer monitor was placed behind the board to be used as the background heat source.

#### C. Thermal Camera and LiDAR Extrinsic Calibration

For the extrinsic calibration between the thermal cameras and the LiDAR, a target-based method may be limited because of the blurs in the thermal images. Instead, we performed a targetless extrinsic calibration.

For the extrinsic calibration between thermal cameras and LiDAR, we defined three costs and combined them with weights for the final optimization. For the first cost, we used a mutual information (MI) between a thermal image and a synthetic LiDAR intensity image as in [35].

$$f_{MI} = \sum_Y \sum_X P(X, Y) \log \frac{P(X, Y)}{P(X)P(Y)} \quad (1)$$

, where  $X$  and  $Y$  represent images from heterogeneous domains (e.g., thermal image and LiDAR intensity image).

For the MI cost minimization, we found directly using the raw thermal images would empirically be intractable. The LiDAR and thermal image modality differ, yielding a lower correlation in the histogram distributions of the heterogeneous sensors. Instead, we propose using ground-based MI cost minimization for the thermal-LiDAR extrinsic calibration. In an urban environment, the grounds heats up

<sup>1</sup><https://sites.google.com/view/rpmsthereo/>

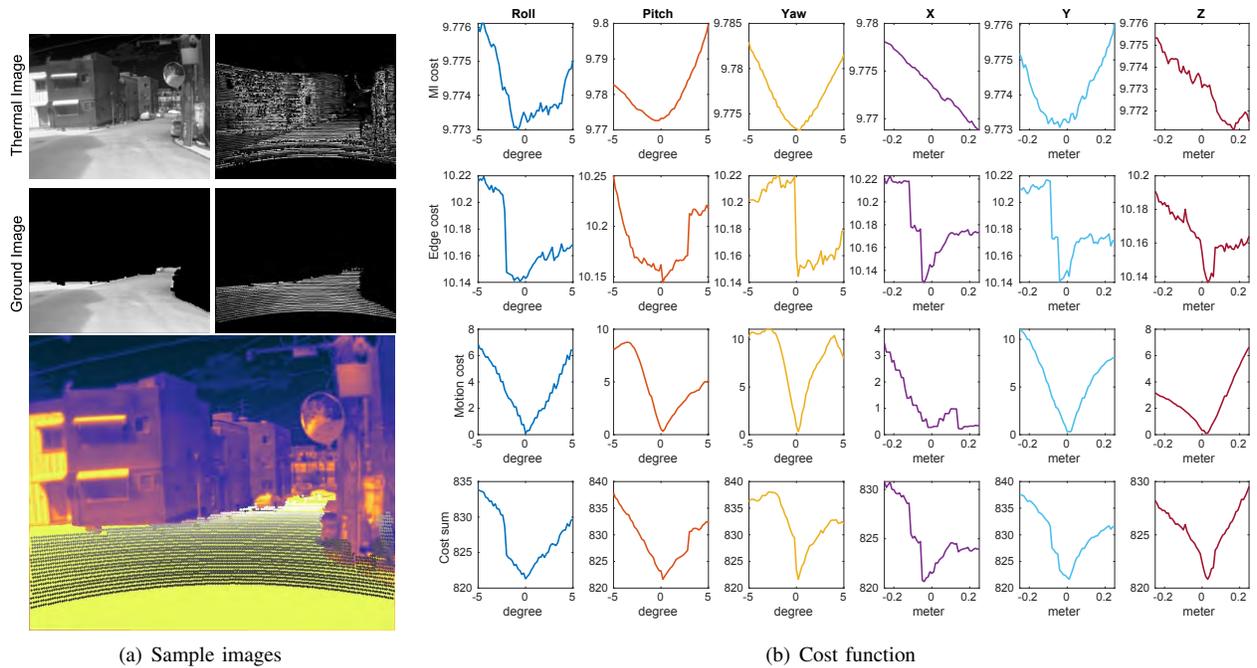


Fig. 3: (a) The top rows are the original and a synthetic thermal images, which has pixel values from the intensities of the projected corresponding LiDAR point cloud of a single scan. The second rows are an automatically ground-segmented thermal image from the original one and the projected LiDAR intensity image from a plane-extracted point cloud. On the bottom, the LiDAR points are well overlaid on the thermal image (the yellow region is the segmented ground). (b) The first three columns in the graph represent the cost by applying rotation perturbation according to roll, pitch, and yaw direction from -5 degrees to 5 degrees. The next three columns are cost graphs representing perturbation for translation from -0.25m to 0.25m along the x, y, and z axes. The rows from top to bottom represent ground MI, edge, motion, and final costs, respectively. The MI cost was effective at constraining pitch, yaw, and y-direction (lateral). The MI cost is relatively weak constrained, while the motion cost successfully constrained all directions except for the x-direction (scene depth). This weak x-directional translation is finally confined by the edge cost. By summing up, the final cost used for optimization shows an optimal graph at 0.

faster than the other objects on a sunny day. Consequently, the ground pixels tend to have uniform and more significant values. By segmenting the ground pixels from the thermal images and extracting a plane from LiDAR, we defined a summed MI cost utilizing a set of multiple pairs from the thermal cameras and the LiDAR sensor. In Fig. 3(a), sample images from each step are visualized.

The second cost aligns the edge of thermal image and LiDAR's pointcloud. Leveraging Gaussian-based edge detection for the thermal image, we apply distance transform to the edge of the thermal image. Similarly, LiDAR's discretized points using depth and intensity value are projected on a thermal image and undergo distanced transformation. Lastly, we apply pixel-wise multiplication between two images.

$$f_{edge} = C - \sum_{u,v} E_C(u,v)E_L(u,v) \quad (2)$$

, where  $E_C$  and  $E_L$  represent the distance transformed thermal and LiDAR's edge images respectively.

The third cost, the motion cost, exploits the relative motion estimation from the LiDAR pair and the thermal camera pair. If the extrinsic calibration is correct, the visual odometry will be aligned to LiDAR odometry. Our approach is to convert

the extrinsic calibration to the visual odometry problem by using the extrinsic parameter as a variable. Accordingly, We employed the relative pose between the previous frame and the current frame of LiDAR through Iterated Closest Point (ICP). After obtaining a relative pose based on the LiDAR, a 3D point is simultaneously projected onto the image plane of  $t-1$  and  $t$  frame of the thermal camera. For these projected points, we measure radiometric error around 3 by 3 patches. The cost is determined by summing these radiometric errors for all projected points between sequential frames of the thermal image and normalized using the number of points,  $N$ .

$$f_{motion} = \frac{\sum I_{t-1}(u_{t-1}, v_{t-1}) - I_t(u_t, v_t))^2}{N} \quad (3)$$

, where  $I_{t-1}$  and  $I_t$  represent the t-1 and t-th camera frames.

#### IV. STEREO THERMAL CAMERA DATASET

##### A. Sequence Summary

The thermal camera measures infrared light, which is related to the temperature of object surfaces. Therefore, we produced multiple sequences from various times and places to gain a diverse dataset. Table IV represents each sequence's

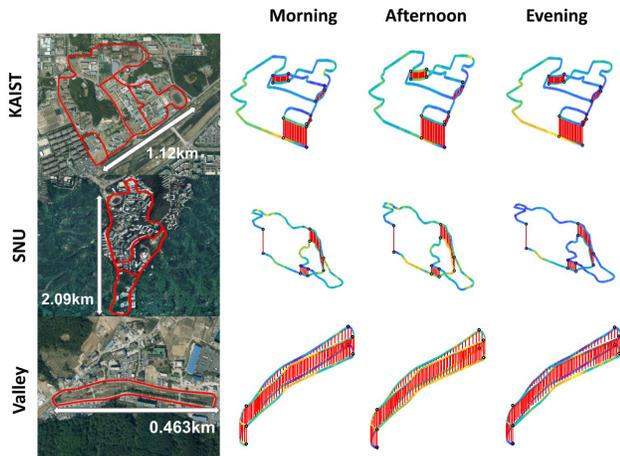


Fig. 4: GT trajectory overlapped with the aerial map. The right image’s color represents the temperature. The closer the blue is to yellow, the higher the temperature. The colormap in each trajectory is independently normalized for clearly visualizing of the temperature differences within a single sequence. We can see more distinct temperature differences in a single trajectory, as observed in the *SNU* and *Valley* sequences because these sites are surrounded by a mountain or the waterside. Also, the start and end point of the loop closure is marked with a black circle, and the loop closure is connected with a red line.

specifications. In our dataset, three different locations are available, and three sequences per location at different times are acquired. The sites’ trajectories and aerial views are visualized in Fig. 4.

1) *KAIST*: *KAIST* is a campus environment with few dynamic objects. These sequences share the same regions of our other dataset, *MulRan* [33].

2) *SNU*: *SNU* is also a campus surrounded by a mountain. Interestingly, the trajectories of *SNU* sequences’ trajectories have a large altitude difference along the mountain between places in them (i.e., maximum 149 m) compared to other sequences of *KAIST* and *Valley*.

3) *Valley*: The *Valley* sequences are located along a waterside. People and moving cars frequently appeared in the *Valley*’s day sequence, compared to the other two sites. Also, the *Valley* sequences are designed to have many loops to support loop detection researches as well as short-term odometry and mapping. The temperature (i.e., thermal intensity) differences are clearly seen in among the first, second, and the third rows in Fig. 6.

Also in Fig. 6, a set of example frames of our dataset is organized. We can see the robustness of thermal images compared to the conventional visual images under illumination changes. We placed a big artificial target (e.g., *ArUco* marker [36], 1000 mm×1000 mm was used) at each site, as can be seen at the beginning and end of a sequence, to support the elaborate evaluations of odometry algorithms.

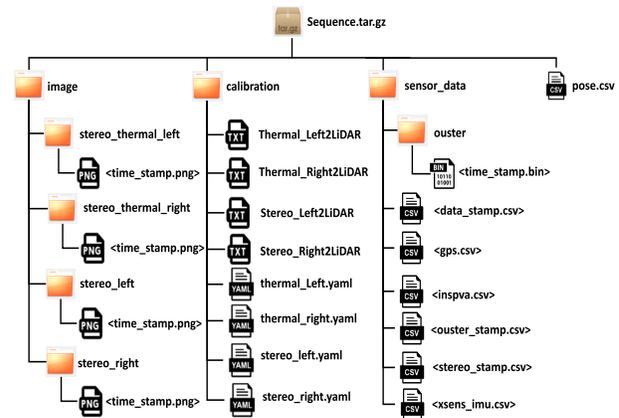


Fig. 5: The directory structure of our dataset.

## B. Sensors and Data Format

We provide the individual sensor data as files to effectively manage the whole dataset’s size and to support easy access to each frame. We also provide Robot Operating System (ROS)-based player, which reads the files and publishes them into ROS topics, to make the real-time SLAM algorithm development easy. Our dataset’s file structure is described in Fig. 5. The details of each sensor data are as follows.

1) *Stereo Thermal Camera Data*: We stored the left and right data of the thermal camera in the form of `<time stamp.png>` in the `image/stereo_thermal_14_left` and `image/stereo_thermal_14_right` paths, respectively. The thermal image shape is  $640 \times 512$  resolution of 10 Hz. We provide 14-bit raw thermal images for various applications rather than 8-bit normalized images, which are susceptible to the normalization range and a scene’s particular temperature. Due to NUC, two thermal cameras run asynchronously. The occurrence of the NUC event is more frequent during the day than night.

2) *Stereo RGB Camera Data*: The RGB camera data are provided in `image/stereo_left` and `image/stereo_right` directories, and are named as `<time stamp.png>`. RGB images are synced in 10 Hz using an external trigger. The RGB image shape is  $1280 \times 560$ . The night (i.e., low light environment) data were acquired by adjusting the gain and exposure time to maintain the image quality as well as the 10 Hz data stream.

3) *3D LiDAR Data*: A set of 3D LiDAR data is saved as a binary file per scan and named as `<time_stamp.bin>` in `sensor_data/ouster` file. Each binary file contains  $x$ ,  $y$  and  $z$ , and intensity information of 1024 width and 128 rays (heights) of a single LiDAR scan.

4) *GPS, INS Data*: The directory `sensor_data/gps.csv` stores the GPS data of Novatel CPT7 INS sensor. The `.csv` file contains the `time_stamp`, latitude, longitude, altitude, and position covariance. The directory `sensor_data/inspva.csv` provides not only GPS information but also information on rotation through correction made via GPS and INS. The member in the file are in the following order: `time_stamp`,

TABLE III: Sequence description of our Dataset

Sequence Name	Location	Time	Description	Path length	Duration
KAIST01	Yuseong, Daejeon	Morning (2021-08-28-08-27-15)	Flat campus Many loops	7.55 km	1084s
KAIST02	Yuseong, Daejeon	Day (2021-08-28-15-28-46)		7.55 km	1351s
KAIST03	Yuseong, Daejeon	Night (2021-08-28-00-51-08)		7.55 km	1130s
SNU01	GwanAk, Seoul	Morning (2021-09-11-08-36-30)	High altitude variance campus High quality of GPS signal	8.75 km	1138s
SNU02	GwanAk, Seoul	Day (2021-09-04-14-18-33)		8.75 km	1246s
SNU03	GwanAk, Seoul	Night (2021-09-01-22-10-45)		8.75 km	1525s
Valley01	Yuseong, Daejeon	Morning (2021-08-28-10-53-37)	Waterfront environment Shortest sequence A lot of dynamic objects	2.08 km	418s
Valley02	Yuseong, Daejeon	Day (2021-08-28-17-58-38)		2.08 km	372s
Valley03	Yuseong, Daejeon	Night (2021-08-27-23-13-46)		2.08 km	355s

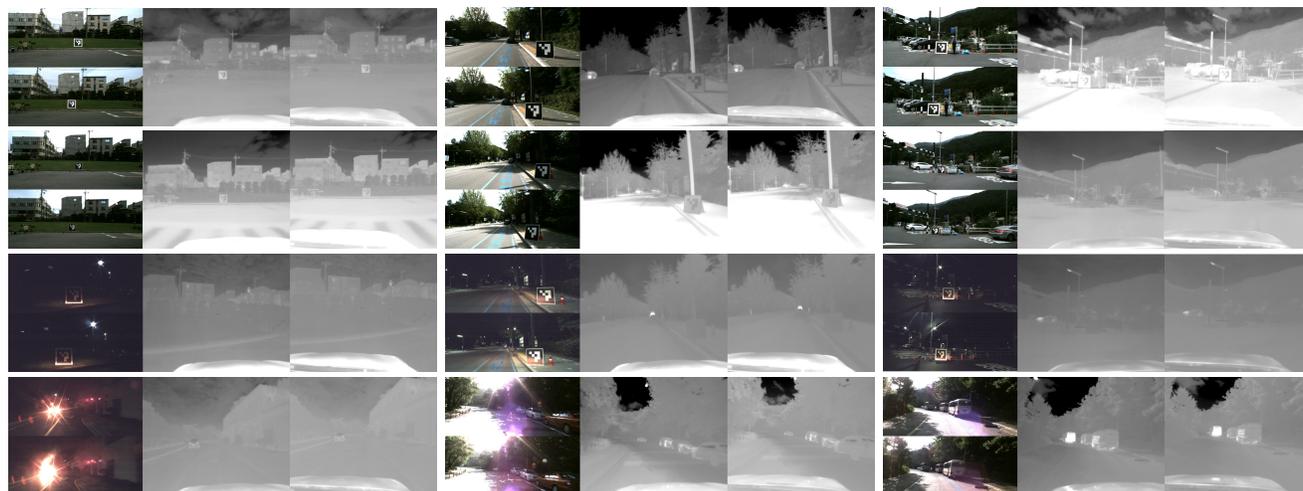


Fig. 6: Examples of stereo thermal frames and corresponding visual images for the scene. The three groups on each column represent the KAIST, SNU, and Valley sequences. The first three rows correspond to morning, day, and night. The last row provides example thermal images' robustness under heavy lights where the conventional visual images were damaged. The thermal cameras are robust against illumination variance, as can be seen in the third and fourth rows.

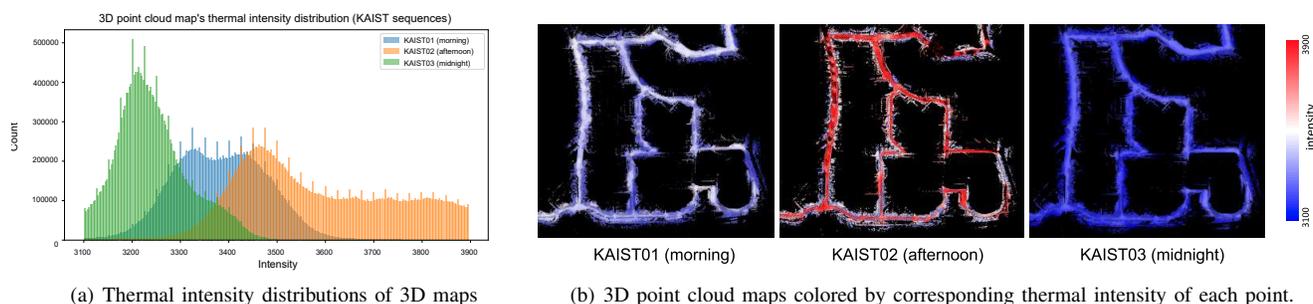


Fig. 7: Our potential application (i.e., thermal-LiDAR mapping). (a) The thermal intensity distribution in histogram. This histogram shows the temperature distribution that changes with respect to the data acquisition time. The temperature in the afternoon reveals a higher average temperature than others. (b) 3D point cloud maps for the same site (i.e., KAIST) with respect to different timestamps (i.e., morning, day, and night). The intensity value used for the map colors is the 14 bit value from raw thermal measurements.

latitude, longitude, height, north\_velocity, east\_velocity, up\_velocity, roll, pitch, azimuth, and data\_status.

5) *IMU Data*: The sensor\_data/imu.csv file, logged in 100 Hz, contains the following information in order: time\_stamp, quaternion x, quaternion y, quaternion z, quaternion w, euler x, euler y, euler z, gyro x, gyro y gyro z, acceleration x, acceleration y, acceleration z, magnetic-field x, magnetic-field y, and magnetic-field z data.

### C. Ground Truth Trajectory

We provide baseline SE(3) trajectories expected as a robot's ground truth poses in UTM global coordinates. The LiDAR is the body frame of our dataset and the baselines are generated via LiO-SAM [37] with RTK-GPS. Because RTK-GPS may be sporadically unavailable within a sequence, we excluded uncertain RTK-GPS values when constructing the

TABLE IV: Trajectory estimation using ORB-SLAM2 and DSO over day (Valley01) and night (Valley03) sequences. Mono and stereo images of thermal and RGB cameras were tested. In the table, ‘M’ and ‘S’ represent mono and stereo camera. Translation and rotation were evaluated through RE and ATE. RE translation and rotation represent interval error rate and degree per 10m. ATE means RMSE for the entire path. If tracking loss occurred during the evaluation process, it was marked with ‘—’.

Sequence Name	Sensor type	ORB				DSO			
		trans (%) (RE)	rot (RE) [°/10m]	trans (AE[m])	rot (AE[°])	trans (%) (RE)	rot (RE)[°/10m]	trans (AE[m])	rot (AE[°])
Valley01	Thermal (M)	1.97	0.08	34.11	17.37	11.66	0.26	132.03	15.38
	RGB (M)	0.71	0.02	9.82	2.41	10.43	0.12	55.45	7.7
	Thermal (S)	1.31	0.13	15.02	12.03	11.1	1.01	127.52	76.71
	RGB (S)	0.37	0.04	6.51	3.91	8.88	0.21	57.38	28.31
Valley03	Thermal (M)	7.82	0.1	87.66	30.43	12.31	0.12	115.65	20.32
	RGB (M)	-	-	-	-	-	-	-	-
	Thermal (S)	5.86	0.38	47.44	23.95	-	-	-	-
	RGB (S)	-	-	-	-	-	-	-	-

baseline trajectory by using covariance of the measurement. Each pose is calculated using LIDAR and IMU factors in the pose-graph model, and correction is performed using RTK-GPS. After that, it is converted into UTM global coordinate using GPS information. All of the pose using this method are stored in sensor\_data/pose.csv .

#### D. SLAM evaluation on stereo thermal dataset

Utilizing the baseline trajectory of our dataset, the two most usual methods of visual SLAM were evaluated. For this evaluation, the representative method of feature-based approach (ORB-SLAM2 [14]) and the direct-based approach (DSO [38]) were used. The estimation accuracy was evaluated for the thermal cameras (Mono/Stereo) and RGB cameras (Mono/Stereo). Since ORB-SLAM and DSO use 8-bit images, the evaluation was performed by normalizing the 14-bit thermal image to an 8-bit image through constant min and max values. For evaluation metric, RPG trajectory evaluator [39] was utilized by using baseline trajectory produced via LiDAR SLAM and RTK-GPS.

As listed in Table IV, both RE and ATE were reported for two sample sequence from morning and night. In the morning with sufficient illumination (Valley01), the odometry from the RGB cameras substantially outperformed the thermal cameras. However, at night when visual degradation occurred (Valley03), the odometry using the RGB camera encountered severe tracking loss, whereas using a thermal camera yielded robust results regardless of the illumination variance.

Notably, large errors for monocular and tracking loss for stereo images were reported for DSO especially when a large rotation motion occurred. The low contrast of the thermal image at night severely deteriorates the estimation performance. In the Valley03 sequence, there is an interval where the mountain occupies most of the image under large rotational motion. Accordingly, the residual pixel difference in the image was not significant, yielding the direct method failure. On the contrary, the features were still available for this texture-less case, enabling ORB-SLAM to outperform DSO. During the evaluation, the 14-bit thermal image was converted to 8-bit by applying the min and max value to minimize data loss using the histogram in Fig. 7(a). During

the conversion, there was a difference in the performance of odometry depending on the value of clipping. We believe the future implementation of 14-bit direct methods could alleviate this conversion issue.

## V. CONCLUSION AND FUTURE WORK

In this paper, we presented the stereo thermal dataset for SLAM researches. We acquired stereo thermal camera sequences as well as multiple navigation sensors such as LiDAR or IMU in environments with various temperatures. The sequences were designed to have enough path length and multiple revisits(e.g., loop-closures) to promote long-term and large-scale SLAM researches across multiple environments. We also provide baseline trajectories for all sequences to make quantitative evaluations possible for odometry and mapping algorithms. We hope that our dataset will boost thermal camera-based SLAM researches for various mobile robot navigation missions, such as odometry, mapping, thermal sensor fusion, and long-term localization. In doing so, we would like to encourage research topics utilizing multiple and potentially asynchronous thermal cameras to complement each other for robust thermal SLAM.

## REFERENCES

- [1] S. S. Shivakumar, N. Rodrigues, A. Zhou, I. D. Miller, V. Kumar, and C. J. Taylor, “PST900: RGB-thermal calibration, dataset and segmentation network,” in *Proc. IEEE Intl. Conf. on Robot. and Automat.*, 2020, pp. 9441–9447.
- [2] J. Jeong, Y. Cho, Y.-S. Shin, H. Roh, and A. Kim, “Complex Urban Dataset with Multi-level Sensors from Highly Diverse Urban Environments,” *Intl. J. of Robot. Research*, vol. 38, no. 6, pp. 642–657, 2019.
- [3] A. Geiger, P. Lenz, and R. Urtasun, “Are we ready for autonomous driving? the KITTI vision benchmark suite,” in *Proc. IEEE Conf. on Comput. Vision and Pattern Recog.*, 2012, pp. 3354–3361.
- [4] A. J. Lee, Y. Cho, S. Yoon, Y. Shin, and A. Kim, “ViViD : Vision for Visibility Dataset,” in *ICRA Workshop on Dataset Generation and Benchmarking of SLAM Algorithms for Robotics and VR/AR*, Montreal, May. 2019, best paper award.
- [5] J. G. Rogers, J. M. Gregory, J. Fink, and E. Stump, “Test your slam! the sub-tunnel dataset and metric for mapping,” in *Proc. IEEE Intl. Conf. on Robot. and Automat.*, 2020, pp. 955–961.

- [6] W. Treible, P. Saponaro, S. Sorensen, A. Kolagunda, M. O'Neal, B. Phelan, K. Sherbondy, and C. Kambhamettu, "Cats: A color and thermal stereo benchmark," in *Proc. IEEE Conf. on Comput. Vision and Pattern Recog.*, July 2017.
- [7] 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 Trans. Intell. Transport. Sys.*, vol. 19, no. 3, pp. 934–948, 2018.
- [8] 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 Trans. Robot.*, vol. 32, no. 6, pp. 1309–1332, 2016.
- [9] Y.-S. Shin and A. Kim, "Sparse depth enhanced direct thermal-infrared SLAM beyond the visible spectrum," *IEEE Robot. and Automat. Lett.*, vol. 4, no. 3, pp. 2918–2925, 2019.
- [10] M. R. U. Saputra, P. P. de Gusmao, C. X. Lu, Y. Almalioglu, S. Rosa, C. Chen, J. Wahlström, W. Wang, A. Markham, and N. Trigoni, "DeepTio: A deep thermal-inertial odometry with visual hallucination," *IEEE Robot. and Automat. Lett.*, vol. 5, no. 2, pp. 1672–1679, 2020.
- [11] S. Zhao, P. Wang, H. Zhang, Z. Fang, and S. Scherer, "Tp-tio: A robust thermal-inertial odometry with deep thermalpoint," in *Proc. IEEE/RSJ Intl. Conf. on Intell. Robots and Sys.*, 2020, pp. 4505–4512.
- [12] Y. Lu and G. Lu, "Superthermal: Matching thermal as visible through thermal feature exploration," *IEEE Robot. and Automat. Lett.*, vol. 6, no. 2, pp. 2690–2697, 2021.
- [13] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "Orb-slam: A versatile and accurate monocular slam system," *IEEE Trans. Robot.*, vol. 31, no. 5, pp. 1147–1163, 2015.
- [14] R. Mur-Artal and J. D. Tardós, "Orb-SLAM2: An open-source slam system for monocular, stereo, and rgb-d cameras," *IEEE Trans. Robot.*, vol. 33, no. 5, pp. 1255–1262, 2017.
- [15] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. Montiel, and J. D. Tardós, "ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial, and Multimap SLAM," *IEEE Trans. Robot.*, 2021.
- [16] J. Engel, V. Koltun, and D. Cremers, "Direct sparse odometry," *IEEE Trans. Pattern Analysis and Machine Intell.*, vol. 40, no. 3, pp. 611–625, 2017.
- [17] S. Khattak, C. Papachristos, and K. Alexis, "Keyframe-based Direct Thermal-Inertial Odometry," in *Proc. IEEE Intl. Conf. on Robot. and Automat.*, 2019, pp. 3563–3569.
- [18] J. Zhang and S. Singh, "Loam: Lidar odometry and mapping in real-time," in *Proc. Robot.: Science & Sys. Conf.*, vol. 2, no. 9, 2014.
- [19] —, "Laser-visual-inertial odometry and mapping with high robustness and low drift," *J. of Field Robot.*, vol. 35, no. 8, pp. 1242–1264, 2018.
- [20] Y.-S. Shin, Y. S. Park, and A. Kim, "Dvl-slam: sparse depth enhanced direct visual-lidar slam," *Autonomous Robots*, vol. 44, no. 2, pp. 115–130, 2020.
- [21] S. N. Torres and M. M. Hayat, "Kalman filtering for adaptive nonuniformity correction in infrared focal-plane arrays," *J. of the Optical Society of America A*, vol. 20, no. 3, pp. 470–480, 2003.
- [22] A. J. Yang, C. Cui, I. A. Brsan, R. Urtaşun, and S. Wang, "Asynchronous multi-view SLAM," in *Proc. IEEE Intl. Conf. on Robot. and Automat.*, 2021, pp. 5669–5676.
- [23] T. Mouats, N. Aouf, L. Chermak, and M. A. Richardson, "Thermal stereo odometry for UAVs," *IEEE Sensors J.*, vol. 15, no. 11, pp. 6335–6347, 2015.
- [24] P. Saponaro, S. Sorensen, S. Rhein, and C. Kambhamettu, "Improving calibration of thermal stereo cameras using heated calibration board," in *Proc. Intl. Conf. on Image Processing*, 2015, pp. 4718–4722.
- [25] S. Vidas, R. Lakemond, S. Denman, C. Fookes, S. Sridharan, and T. Wark, "A mask-based approach for the geometric calibration of thermal-infrared cameras," *IEEE Trans. Instrum. and Meas.*, vol. 61, no. 6, pp. 1625–1635, 2012.
- [26] P. V. K. Borges and S. Vidas, "Practical infrared visual odometry," *IEEE Trans. Intell. Transport. Sys.*, vol. 17, no. 8, pp. 2205–2213, 2016.
- [27] S. Hwang, J. Park, N. Kim, Y. Choi, and I. So Kweon, "Multispectral pedestrian detection: Benchmark dataset and baseline," in *Proc. IEEE Conf. on Comput. Vision and Pattern Recog.*, June 2015.
- [28] N. Carlevaris-Bianco, A. K. Ushani, and R. M. Eustice, "University of michigan north campus long-term vision and lidar dataset," *Intl. J. of Robot. Research*, vol. 35, no. 9, pp. 1023–1035, 2016.
- [29] M. Burri, J. Nikolic, P. Gohl, T. Schneider, J. Rehder, S. Omari, M. W. Achtelik, and R. Siegwart, "The EuRoC micro aerial vehicle datasets," *Intl. J. of Robot. Research*, vol. 35, no. 10, pp. 1157–1163, 2016.
- [30] W. Maddern, G. Pascoe, C. Linegar, and P. Newman, "1 year, 1000 km: The oxford robotcar dataset," vol. 36, no. 1, pp. 3–15, 2017.
- [31] D. Schubert, T. Goll, N. Demmel, V. Usenko, J. Stückler, and D. Cremers, "The tum vi benchmark for evaluating visual-inertial odometry," in *Proc. IEEE/RSJ Intl. Conf. on Intell. Robots and Sys.* IEEE, 2018, pp. 1680–1687.
- [32] W. Wang, D. Zhu, X. Wang, Y. Hu, Y. Qiu, C. Wang, Y. Hu, A. Kapoor, and S. Scherer, "Tartanair: A dataset to push the limits of visual slam," in *Proc. IEEE/RSJ Intl. Conf. on Intell. Robots and Sys.* IEEE, 2020, pp. 4909–4916.
- [33] G. Kim, Y. S. Park, Y. Cho, J. Jeong, and A. Kim, "Mulran: Multimodal range dataset for urban place recognition," in *Proc. IEEE Intl. Conf. on Robot. and Automat.* IEEE, 2020, pp. 6246–6253.
- [34] D. Barnes, M. Gadd, P. Murcutt, P. Newman, and I. Posner, "The oxford radar robotcar dataset: A radar extension to the oxford robotcar dataset," in *Proc. IEEE Intl. Conf. on Robot. and Automat.* IEEE, 2020, pp. 6433–6438.
- [35] G. Pandey, J. R. McBride, S. Savarese, and R. M. Eustice, "Automatic extrinsic calibration of vision and lidar by maximizing mutual information," *J. of Field Robot.*, vol. 32, no. 5, pp. 696–722, 2015.
- [36] F. J. Romero-Ramirez, R. Muñoz-Salinas, and R. Medina-Carnicer, "Speeded up detection of squared fiducial markers," *Image and vision Computing*, vol. 76, pp. 38–47, 2018.
- [37] T. Shan, B. Englot, D. Meyers, W. Wang, C. Ratti, and D. Rus, "Lio-sam: Tightly-coupled lidar inertial odometry via smoothing and mapping," in *Proc. IEEE/RSJ Intl. Conf. on Intell. Robots and Sys.*, 2020, pp. 5135–5142.
- [38] J. Engel, V. Koltun, and D. Cremers, "Direct sparse odometry," *IEEE Trans. Pattern Analysis and Machine Intell.*, vol. 40, no. 3, pp. 611–625, 2018.
- [39] Z. Zhang and D. Scaramuzza, "A tutorial on quantitative trajectory evaluation for visual(-inertial) odometry," in *Proc. IEEE/RSJ Intl. Conf. on Intell. Robots and Sys.*, 2018.