All-Day Visual Place Recognition: Benchmark Dataset and Baseline

Yukung Choi Namil Kim Kibaek Park Soonmin Hwang Jae Shin Yoon In So Kweon
Korea Advanced Institute of Science and Technology (KAIST), Republic of Korea
[ykchoi, nikim, kbpark, smhwang, jsyoon]@rcv.kaist.ac.kr, iskweon77@kaist.ac.kr

Abstract

This paper introduces all-day dataset captured from KAIST campus for use in mobile robotics, autonomous driving, and recognition researches. Totally, we captured 42 *km sequences at* 15∼100*Hz using multiple sensor modalities such as fully aligned visible and thermal devices, high resolution stereo visible cameras, and a high accuracy GPS/IMU inertial navigation system. Despites of a particular scenario, we provide the first aligned visible/thermal all-day dataset, including various illumination conditions: day, night, sunset, and sunrise. With this dataset, we introduce multi-spectral loop-detector as a baseline. We will open all calibrated and synchronized datasets*[1]*, and hope to make a various state of the art computer vision and robotics algorithms.*

1. Introduction

Localization and mapping problems have played an important role in the development of navigating the autonomous robot and vehicles such as personal service robots, smart cars, and unmanned aerial systems.

Among several useful sensors such as vision-based (visible/thermal cameras), scan-based (2D/3D Lidar), visible cameras have been extensively exploited for visual localization and mapping communities, due to low cost, power consumption and portability. Actually, the most popular localization systems [4, 6, 5, 13] have mainly adopted the visible-sensors and dealt with many localization challenges in real world environments. Also, most public datasets consist of visible-images.

Despite the successful novelty, there are apparent limitations of visible sensor in physically, which is highly sensitive to the illumination changes and moving object detections. For example, night time driving has a difficulty for forward visibility and conversely, can be disturbed by headlights. Also, object appearances cannot be guaranteed constantly, when it is affected by illumination conditions such as shadow, saturation. Recently, there has been increasing interest in using alternative sensors. Thermal sensors are more robust to illumination changes and well distinguished in the scenario where temperature variations are present. Moreover, the thermal devices have become accessible in general use. However, thermal-spectral images sometimes provide ineffective information without any radiant sources and even it has trouble in conversion phenomenon called crossover, and reflection.

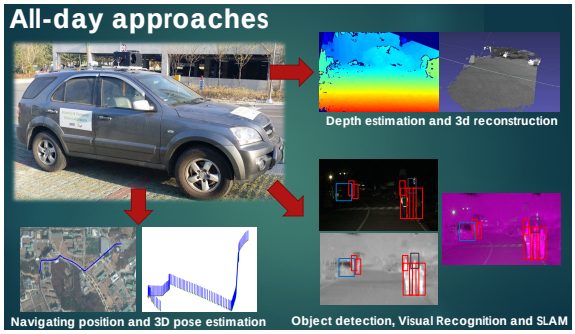


Figure 1. Recording platform with cameras (top-left), trajectory from GPS/IMU units (down-left), depth map and reconstructed 3D points (top-right), and moving object labels (bottom-right).

Other alternative sensors, such as 2D/3D Lidar, can provide useful 3D or depth information of the scene, yet it is not appropriate for practical usage, because of high prices. Therefore, we make the multimodal dataset using visible and thermal spectral sensors. To maximize the complementary information between visible and thermal images, we use the beam-splitter hardware to try to completely align two images [10]. Also, we add a visible-camera for capturing multi-view images and a combined GPS and IMU system for analyzing the physical quantity.

We capture the same loop/place at different times in different conditions. From these efforts, we can produce the aligned multimodal dataset including all-day conditions. We believe that our configuration system put together many challenging issues as above mentioned with combination of complementary information. Furthermore, we expect our dataset to push forward the progress of not visual place recognition, also further extension of computer vision and robotics algorithms.

[1] All-day visual place recognition dataset is available online: http://rcv.kaist.ac.kr/all-day/

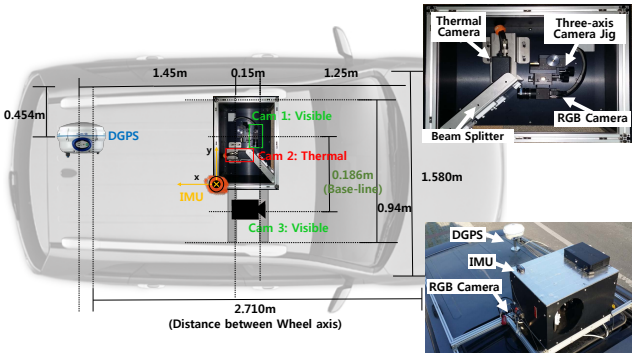


Figure 2. **Hardware configuration**. (Left) Capturing platform, (Top) Beam-splitter, (Bottom) Side View of Sensor setup.

2. Configuration and Methodology

2.1. Sensors and Data Acquisition

Sensors We equipped multimodal sensors onto a standard sport utility vehicle illustrated in Fig. 2.

2 $\times$ PointGrey Flea3 color camera(FL3-GE-13S2C-C), $30Hz$, 1.3 Megapixels, $1/3$" Sony ICX445 CCD, 1280×960, Global shutter, $400 \sim 750$nm, GigE

2 $\times$ Computar Optics $12mm$ Lens, $26°$(H) $\times$ $22.1°$(V)

1 $\times$ FLIR A35 thermal camera, $60Hz$, $1.32mrad$, 320×256, $25°$ (H) $\times$ $19°$ (V) with $19mm$ lens, $7.5 \sim 13$um, GigE

1 $\times$ Trimble GPS with RTK (open sky localization errors $< 5cm$), $10Hz$ and MicroStrain 3DM-GX3 IMU unit (Resolution $< 0.1°$), $100Hz$

Note that the visible and thermal sensors are combined with beam-splitter, which is made of zinc-oxide and silicon materials. This optical device can reflect visible wavelengths and transmit long-wavelengths infrared lights (LWIR). Using the beam-splitter setup, the alignment of visible and thermal is completely parallax-free and the highly accurate mapping of the lower resolved thermal image in the visible image is easily done without additional use of image rotating or straightening algorithms. The baseline of both cameras is approximately 18.6 cm.

Data Acquisition We mounted our multimodal capturing system on top of vehicle, and housed a PC with Intel i7-4970K processors in the trunk. To prevent the bottleneck of file i/o, we used 3 solid state drivers (SSD) to save each sensor data separately. One SSD saved thermal and GPS/IMU data, and other two SSDs saved two visible sensors.

2.2. Sensor Calibration

For the purpose of our system, all sensors are carefully calibrated and synchronized. Visible-spectral calibration has been researched in many applications, and there are already many reliable methods and toolkits. We used the one of most popular calibration toolkits[2] to automatically calibrate intrinsic and extrinsic parameter of the individual and multiple visible cameras. However, due to the different spectral range, a joint calibration between visible and thermal is not easy task for obtaining accurate results. Therefore, we present an effective approach for geometrically calibrating visible and thermal sensors using a new calibration pattern board in Fig. 3.

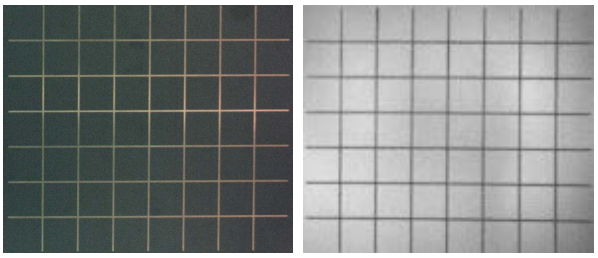


Figure 3. **Proposed calibration board**. (Left) Visible imaging, (Right) Thermal imaging

Multi-spectral calibration For the thermal-spectral imaging, popular chessboard approaches involve heating the pattern through exposure to a flood lamp [3, 17]. Recently, Hilsenstein [9] manufactured the chessboard pattern onto a specially manufactured printed circuit board (PCB), and Vidas et al. [20] proposed a simple calibration framework using affordable materials by hand-made. Despite all these efforts, the corner points cannot be reliably extracted in a thermal-image, because, physically, the spread of thermal is proportional to the area of pattern.

We took account of a limitation and designed a calibration board consisting of line-grids of regular sized squares (Fig. 3). The thin cooper line is milled onto a printed circuit board at equal spaces. Lines were 2 *mm* wide and spaced with 40 *mm* separation, and a six intersections along the shorter axis and seven intersections along the longer axis were used. These dimensions are selected by two folds:

(1) A geometric correspondence with chessboards

(2) Field of view and resolution of thermal camera

A significant importance of this board is that it can achieve better results involving a high contrast at thermal-domain due to the uniform temperature. Also, the proposed calibration board can be compatible with many existing camera toolkits, due to the same geometric relations to the chessboard pattern calibration. Therefore, we performed calibration using popular camera calibration toolbox with our pattern board in the same way. To align both cameras, we calibrated the intrinsic and extrinsic parameters of the cameras, and controlled the position until a relative rotation and translation are a negligible.

[2]Caltech calibration toolkits: http://www.vision.caltech.edu/bouguetj/calib_doc/

2.3. Synchronization

The popular approach is Precision Time Protocol (PTP) synchronization, which uses ethernet time throughout a sensor network. However, since all sensors should share a same ethernet adapter, there is a obvious risk for camera network bandwidth. For our system, PTP approach is not adequate for handling high resolution images and multiple sensors.

Therefore, we adopted high precision synchronization with respect to the external trigger. Since the thermal camera provides a external sync channel, we set the thermal camera as a master and two visible cameras as the slave. The details are given as

Trigger generation (Master: FLIR A35)

$*$ Trigger signal : $3.3V$ pulse with $29.97Hz$

$*$ Register/PIN : *SelfSyncMster* / PIN 3 $(SYNC\_OUT)$

Trigger receiver (Slave : FL3-GE-13S2C-C)

$*$ Mode : $OverlappedExposure/Readout(Mode14)$[3]

Unfortunately, the GPS/IMU system cannot be synchronized in this way, as they does not support external triggering. Instead, we collected the closest timestamp to the camera timestamp for a particular frames. All timestamps have been recorded on our computer using the system clock.

3. Data Collection

Our all-day dataset was captured around the Korea Advanced Institute of Science and Technology (KAIST) campus. Despite of a particular space, it is a novel dataset including a diverse illumination changes and dynamic objects, and having a potentiality to deal with these challenges with the aligned multimodal images (visible and thermal), multi-view images (stereo) and GPS/IMU data. The data was collected while driving the vehicle around the same environment in different times. Specially, our dataset involves not only day and night-time but the ambiguous time such as sunset and sunrise. In each time, we took care of checking over all sensor configurations and calibration errors by a simple test, and covering many cases in real world environments. A sample trajectory of the dataset in one loop around the campus is shown in Fig. 5.

3.1. Data Description

The unprocessed data consists of five files for each trial. Three files contain raw images of two visible and one thermal image captured at 15 fps. Other files contain the data from remaining sensors; GPS and IMU at 10fps and 100 fps respectively. This file stores the data in the format similar to that described in KITTI [7], which is one of famous datasets for visual place recognition. The unprocessed data consists of raw distorted images from visible and thermal camera and the raw navigating/inertial data.

[3]FLIR does not recommend using the external sync interface with a slow-configured camera because there is ambiguity as to which received pulse triggers the frame timing.

Figure 5. **GPS Trajectory for one sequence**. Here we plot the GPS overlaid ontop of an aerial image from Google maps. Colors encode the GPS signal quality: blue tracks have been recorded with high precision using RTK corrections, green denotes the general GPS information without correction signals. The red indicates the shadow area which one have been excluded from our data set. W, E and N indicate each sequence in our datasets.

Images To obtain various illumination changes, we captured the same place successively in fixed time (4:00, 6:00, 11:00, 14:00, 18:00, 24:00). The time considers the elevation angle of the sun including sunrise, sunset, dawn, midday and night. Therefore, our dataset can cover all-day natural illumination changes and be possible to analyze on how the light conditions affects the multispectral imaging. One example of dataset are shown in Fig. 4. We show that each spectral image shows the different tendency according to the amount of radiation. For example, at 18:00 after sunset, thermal images maintain a high contrast as the radiation is decreasing. On the contrary, at 6:00 after sunrise, thermal images are still low intensity distributions.

Train and Test sets. Instead of randomly dividing, our dataset can be categorized into three groups according to the light condition. One is day-time (11:00, 14:00) groups, and another is night-time groups (24:00, 04:00). The other group consists of the time (6:00, 18:00) at dawn and dusk, when the light extremely changes.

Moving object annotation. Moving object detection is one of challenging problems in the place recognition problems. To handle this problem, we manually annotated the subset for targeting some objects such as pedestrians and vehicles. When we annotated the ground truth, we used the modified Piotr's Computer vision MATLAB Toolbox [10] which is adapted for displaying visible and thermal images simultaneously.

Names	#Sequences/#category	Total Length (km)	#Frames	Image size	Average Speed (km/h)	Image type	FOV(°H×°V)	Description	Condition	GPS	Multi-spectral	Publish	Dataset Open
New College [18]	1	2.2	51k	0.2M 384×512	5.4	visible	96°×86°	Outdoor Robot	Sunny	✓	-	'09	✓
Malaga 2009 [1]	2	6.4	38k	0.8M 1024×768	10.1	visible	-	Outdoor Vehicle	Sunny	✓	-	'09	✓
Ford Campus [16]	2	5.1	7k	1.0M 1600×600	56.2	visible	omni	Outdoor Vehicle	Sunny	✓	-	'10	✓
KITTI Dataset [7]	22 5	39.2	41k	0.8M 1385×512	18	visible	90°×35°	Outdoor Vehicle	Sunny	✓	-	'12	✓
St Lucia MTD [8]	10	-	-	- 640×480	-	visible	62°×48°	Outdoor Vehicle	Illumination change	✓	-	'10	✓
Alderley [13]	2	8.0	32k	0.2M 640×260	52.5	visible	-	Outdoor Vehicle	Sunny Rainy	✓	-	'12	✓
DIRD [11]	3	10.0	7.8k	0.5M 1226×387	-	grey	-	Outdoor Vehicle	Illumination change	✓	-	'13	✓
TUM RGB-D [19]	27	0.4	65k	0.3M 640×480	-	visible depth	57°×43°	Indoor Robot	-	✓*	✓	'11	✓
MSLAM [21]	7 2	-	-	-/14bit (T) 640×480	-	thermal	-	In/Outdoor Bicycle	Illumination change	✓	-	'12	-
MO [14]	5 2	1.8	-	-/16bit (V) 658×492 (T) 640×480	-	visible thermal	-	Outdoor Vehicle	Illumination change	✓	✓	'14	-
SC [2]	3 2	6	-	-/16bit (V) 780×580 (T) 640×480	-	visible thermal	-	Outdoor Vehicle	Moving object	✓	✓	'15	-
QUT MSLAM [12]	10	1.5	6k	1.4M/- (V) 640×480 (T) 640×480	-	visible thermal	-	Outdoor Bicycle	Illumination change	✓	✓	'12	-
Ours	3	42	105k	1.3M/8bit (V) 1280×960 (T) 320×256	24.5	visible thermal	26°×22° 25°×19°	Outdoor Vehicle	Illumination change	✓	✓	'15	✓

Table 1. Comparison of several place recognition datasets. − denotes unmentioned information in reference papers. ∗ denotes a virtual ground truth instead of GPS. This measurement is estimated by external camera. k indicates the 10^3. **Note that our dataset is only one which provides aligned visible-thermal images and moving object labels captured in dynamic environments**.

3.2. Comparision Existing Dataset

We sum up the existing visual localization dataset in Table.1 for the comparison. Many datasets [18, 1, 16, 7] provide visible image sequences captured in day-time, under fine weather condition without extreme illumination changes. Some popular datasets [1, 16, 7] contain the real driving scenarios. Even if Ford [16] dataset relatively gives the smaller number of images, they have a novelty with combining high resolution Omni-directional camera and 3d Lidar. KITTI [7] play a role as a benchmark in various computer vision applications such as stereo vision, optical flow, visual SLAM, and object detection.

Recently, St Lucia MTD, DIRD and Alderley [8, 11, 13] focus mainly on variant of illumination changes. St Lucia MTD [8] was captured at five different times (8:25/10:00/12:10/14:10/15:45) during the day to obtain the appearance variation as time passes. They had taken images over 3 weeks and made total 10 datasets. DIRD [11] provides the dataset including severe changes of illumination and cast shadows over a half-day. Alderley [13] dataset has two different scenarios; a sunny day and a rainy night with GPS positions as ground truth, available for validating from eyes.

There are some researches to handle the night-time place recognition using visible and thermal images. Unfortunately, there is no public dataset. MO [14] consists of visible and thermal images, and stereo images. SC [2] is for segmentation with visible and thermal stereo images. QUT-MSLAM [12] has a similar concept in terms of multimodal system and applications. The dataset is captured by the

Figure 4. These figure are dataset examples captured from the same location during a day. The illumination changes of visible-thermal images are clearly appeared across 24hours.

platform, which is made of visible camera, thermal camera and GPS receiver, attached to the bicycle, in different times of day. Another multimodal dataset is TUM-RGBD [19]. However, the dataset used a Microsoft Kinect sensor and even it is for indoor SLAM.

Even if our datasets belong to multimodal dataset, we have a novelty to provide aligned visible and thermal images during all-day. Compared to QUT-MSLAM [12], our dataset has ego-centric moving view in real-driving conditions, and a pair of high resolution visible images.

4. Dataset Statistics

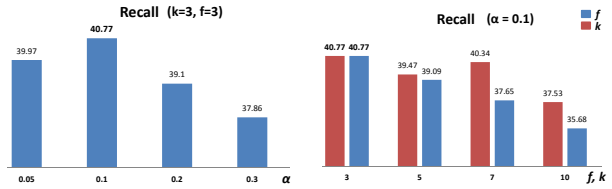
To give further insights into the properties of our dataset, we provide statistics for all sequences in Fig. 6. As above mentioned, moving object, illumination changes and viewpoint changes can cause mis-matching of visual places recognition. In these perspectives, we expect that those types of analytic information help the finding reason why the place recognition is failed. The ratio of moving object and view-point changes can be estimated by annotation and ground truth respectively, and the degree of illumination change can be indirectly predicted by saturation map in the image.

As shown in Fig. 6 (a-d), even if both types of image lose information at night, thermal-images have a smaller ratio of loss than visible-images. Thus, we can conclude that thermal-images assist visible images with complementary information at night. In Fig. 6 (e-i) illustrates the distribution of extracted features in the scene and we visualize the results through the weighted Gaussian histogram. Since most of features are concentrated on an upper band across the image center, we can estimate that the field of view of our system is set as similar as the general driver. The saturation map of visible-images are depicted in Fig. 6 (f-l). After sunset (18:00), there are many under-saturated regions. While the sun rises, we can easily detect over-saturated regions in an upper region of images such as sky. This is an important measurement to determine on how much illumination changes as time passes and how much illumination affects intensities in particular times. Fig. 7 (a-b) shows a statistical analysis for ego-motion of our configuration system for the whole dataset. Despite of many speed bumps in our campus, we included all sequences for reasonable statistics to represent natural driving behavior. The right figures (Fig. 7 (c-d)) indicate the amount of changing angular ratio per the number of images in log-scales. If the angular rate is a high, the view is changed rapidly. We expect these statistics to be helpful for analyzing failure cases in visual place recognition methods employing the proposed dataset.

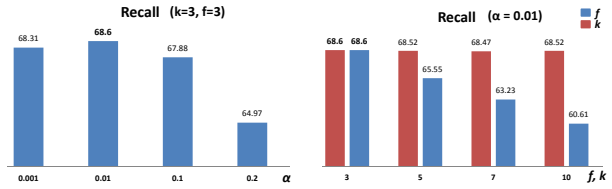
5. Baseline Approach

In this paper, we utilized DLoop [6] as our baseline algorithm to verify the compatibility of previous algorithms and handle the visible and thermal images. This is natural choice because the algorithm is generally used as fast and efficient method for visual place recognition using bag of words and especially considers the temporal and geometric consistency to verify the loop-closing.

We used SURF as the local feature in visible and thermal images. For the extension, we built visible and thermal vocabulary tree [15], respectively. We decided a correct location as any spectral algorithm completed the loop-closing in each frame. In our baseline, we did not use concatenated



(a) Results of visible dataset



(b) Results of thermal dataset

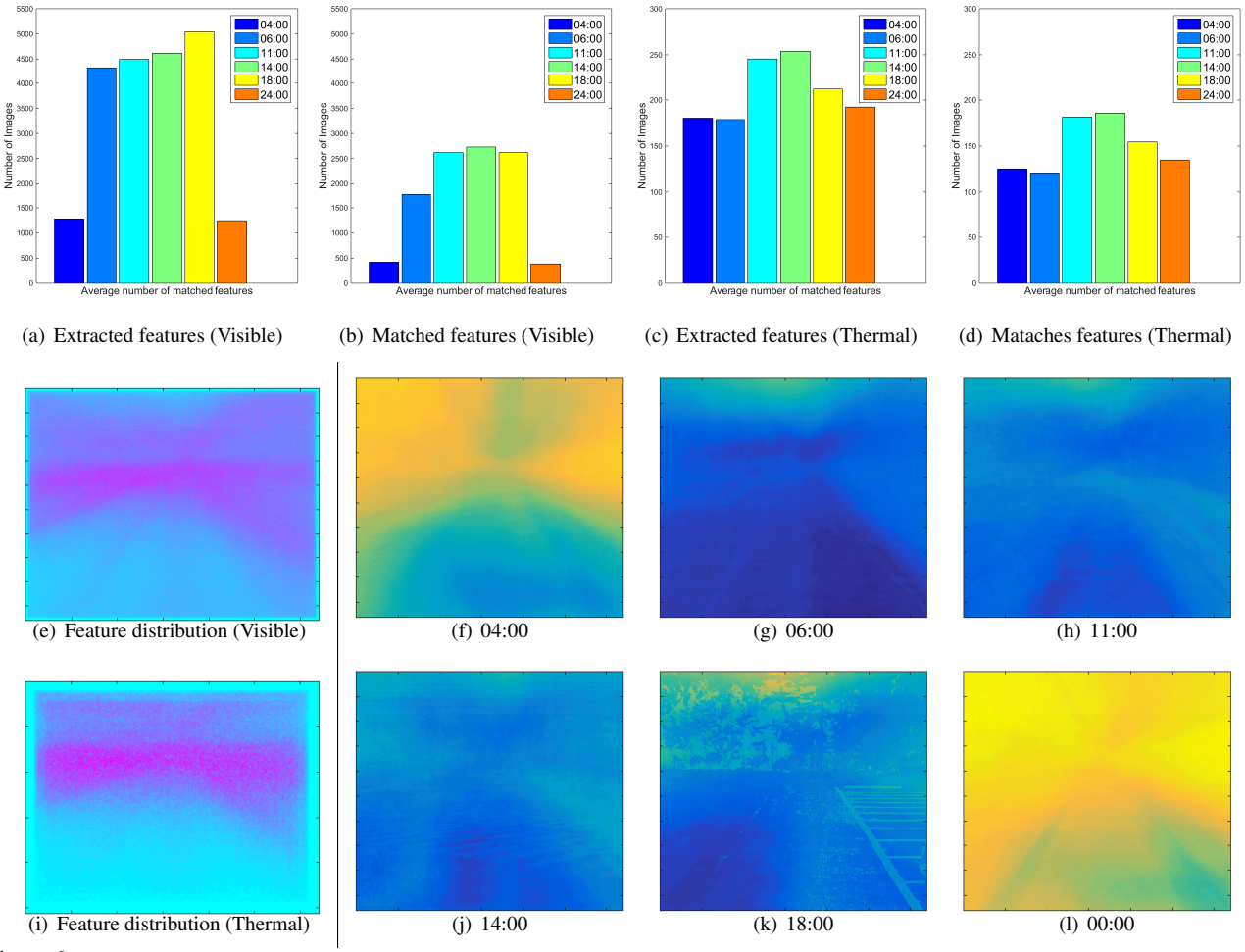
Figure 8. Recall graph in visible and thermal dataset according to parameters. α is similarity threshold, k is related to the temporal consistency, and f is sampling step of train and test dataset. Finally, we chose $\alpha = 0.1$(V), 0.01(T), $k = 3$, and $f = 3$.

descriptor for integrating vocabulary tree, because the different characteristics of visible and thermal images made harder to verify results in DLoop.

For reasonable analyses, we decided proper baseline parameters with regard to original algorithm [6]. Similarity threshold α, temporal consistency k and sampling step f are generally chosen in experimental results due to their dependency on datasets. The optimal parameters of our dataset are shown in Fig. 4. Fig 8 shows the evaluation results of baselines for our dataset. The upper table indicates numerical results of recall in each sequence (West, East, and North). Fig. 9 (a-c) are loop-closing results of visible data and (d-f) are those of thermal data in all sequences. As numerical and loop-closing results, thermal image seems to be dominant at night-time, when scenes are hardly distinguishable in the visible images. However, visible image performed better than thermal image at day-time, when thermal scenes partly suffer from the artifacts such as crossover. The results of extended baseline using both images are shown in Fig. 9 (g-i). Compared with a single-spectral result, our extended baseline acceptably shows better performances in all-times. This implies that the complementary information of visible and thermal image is helpful regardless of any illumination conditions.

6. Summary and Future Work

In this paper, we have presented a well aligned, calibrated and synchronized visual place recognition dataset for all-day long. We also examine our extention of DLoop, called multi-spectral DLoop in various condition. Through those results, we determined that the complementary information of visible and thermal image is helpful in visual place recognition problems. We believe that our dataset

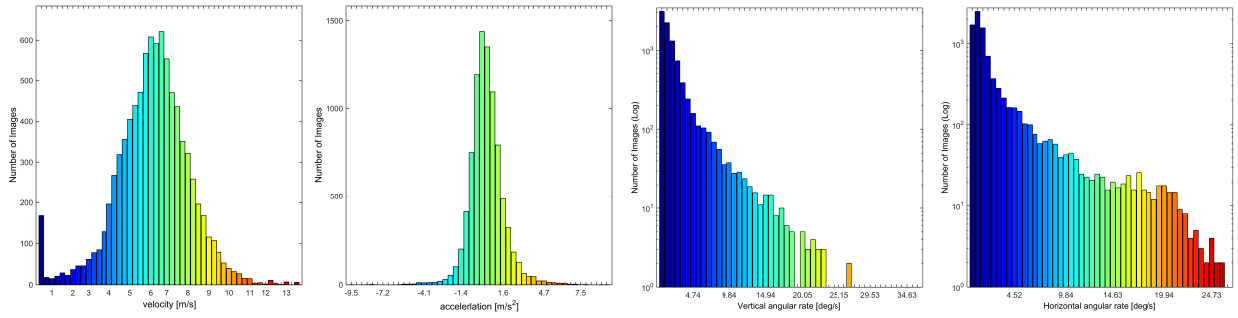


(a) Extracted features (Visible) (b) Matched features (Visible) (c) Extracted features (Thermal) (d) Mataches features (Thermal)

(e) Feature distribution (Visible) (f) 04:00 (g) 06:00 (h) 11:00

(i) Feature distribution (Thermal) (j) 14:00 (k) 18:00 (l) 00:00

Figure 6. Statistics for images. We use the SURF features in OpenCV to reveal the image conditions for visual place recognition. (a-d) The features both on the visible and thermal images are extracted from useful structures on the side of the road. (e,i) We verify that the visible images are easily saturated due to the lack of illumination. (f-h, j-l) These information could be useful for visual place recognition task.

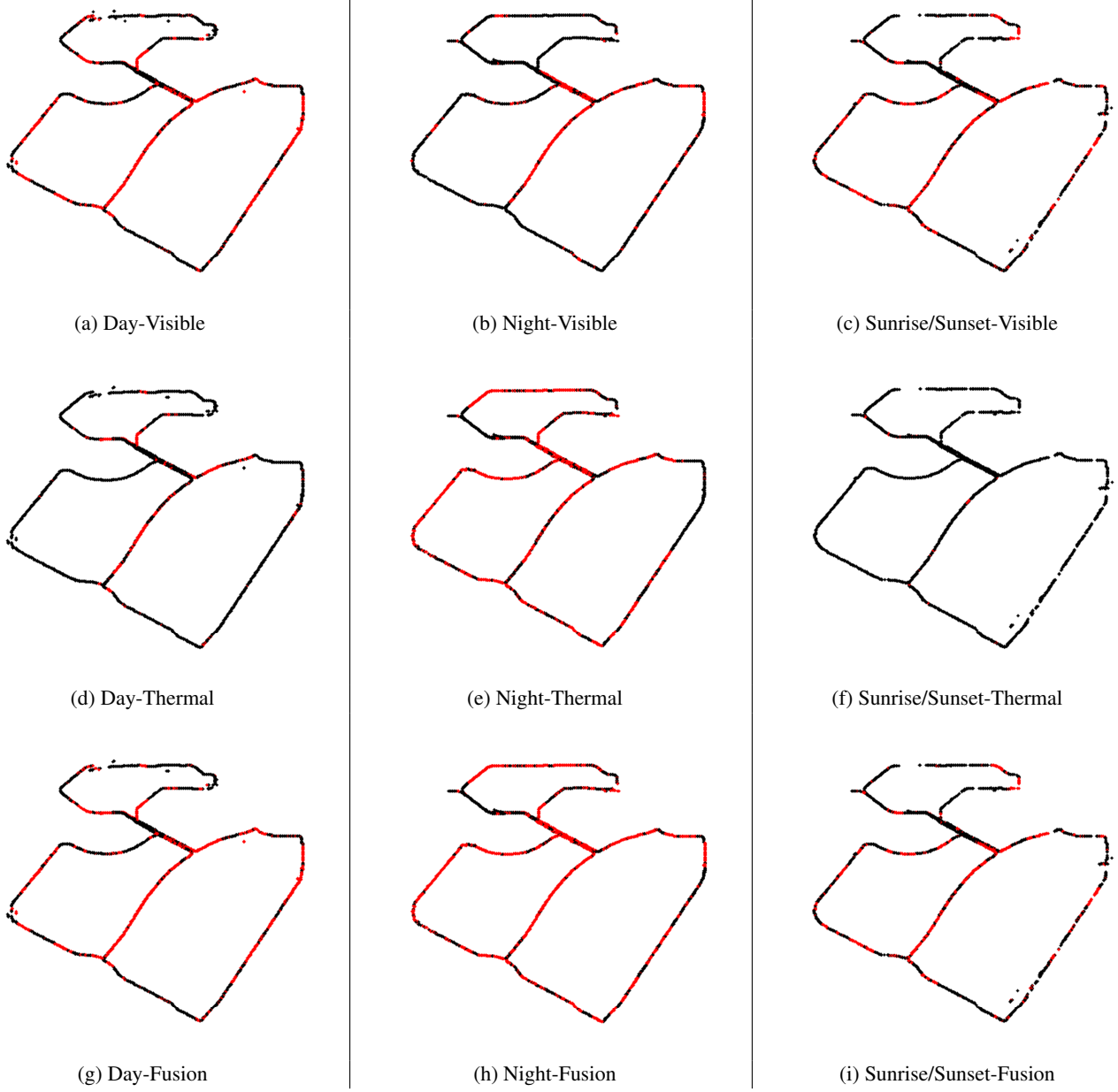


(a) Velocity (b) Acceleration (c) Vertical angular rate (d) Horizontal angular rate

Figure 7. Statistics for egomotion. When we capture the proposed dataset, we try to keep constant velocity. Since there are many speed bumps in our campus, we should repeatedly accelerate and decelerate. (a,b) The proposed dataset also gives the vertical and horizontal angular rate of the scene, i.e. how the scenes are changed rapidly. (c,d) We hope these information might be helpful for analyzing failure case of the algorithms.

can be used as a benchmark for testing various state of the art computer vision and robotics algorithms such as visual place recognition, 3d reconstruction, moving object detection (pedestrian, vehicle) and autonomous driving research field. In the future, we have a plan to expand the amount of dataset for long-term visual recognition, make diverse categories for object detection and capture a variety of scenarios such as road, residential and downtown area in general traffic scene. In particular, we will soon provide the baseline algorithm in several applications based on our dataset.

West	Recall			East	Recall			North	Recall		
	visible	thermal	fusion		visible	thermal	fusion		visible	thermal	fusion
day	48.60	21.58	55.27	day	47.74	16.98	49.94	day	20.42	10.17	23.91
night	39.27	51.21	63.96	night	40.80	66.55	77.67	night	9.93	45.69	49.10
change	30.36	1.23	30.57	change	51.53	3.13	51.67	change	19.50	6.70	22.24



(a) Day-Visible (b) Night-Visible (c) Sunrise/Sunset-Visible

(d) Day-Thermal (e) Night-Thermal (f) Sunrise/Sunset-Thermal

(g) Day-Fusion (h) Night-Fusion (i) Sunrise/Sunset-Fusion

Figure 9. The results for our all dataset. GPS positions of the images are plotted with a black dot. Wherever an image matches a loop with another image, the position is labelled with a red dot. The precision is 100% in all cases. AM04, AM06 and AM11 sets are used to train a vocabulary tree.

References

[1] J.-L. Blanco, F.-A. Moreno, and J. Gonzalez. A collection of outdoor robotic datasets with centimeter-accuracy ground truth. *Autonomous Robots*, 2009. 4

[2] P. V. K. Borges and P. Moghadam. Combining motion and appearance for scene segmentation. In *icra*, 2014. 4

[3] S. Y. Cheng, S. Park, and M. M. Trivedi. Multiperspective thermal ir and video arrays for 3d body tracking and driver activity analysis. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2

[4] M. Cummins and P. Newman. Fab-map: Probabilistic localization and mapping in the space of appearance. *The International Journal of Robotics Research*, 2008. 1

[5] 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 (PAMI)*, 2007. 1

[6] D. Galvez-Lopez and J. D. Tardos. Bags of binary words for fast place recognition in image sequences. *IEEE Transactions on Robotics*, 2012. 1, 6

[7] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun. Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research*, 2013. 3, 4

[8] A. Glover, W. Maddern, M. Milford, and G. Wyeth. Fab-map + ratslam: Appearance-based slam for multiple times of day. In *ICRA*, 2010. 4

[9] V. Hilsenstein. Surface reconstruction of water waves using thermographic stereo imaging. In *Image and Vision Computing New Zealand*, 2005. 2

[10] S. Hwang, J. Park, N. Kim, Y. Choi, and I. Kweon. Multispectral pedestrian detection: Benchmark dataset and baseline. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015. 1, 3

[11] H. Lategahn, J. Beck, B. Kitt, and C. Stiller. How to learn an illumination robust image feature for place recognition. In *IEEE Intelligent Vehicles Symposium*, 2013. 4

[12] W. Maddern and S. Vidas. Towards robust night and day place recognition using visible and thermal imaging. *Proceedings of the Robotics: Science and Systems Conference*, 2012. 4, 5

[13] M. J. Milford and G. F. Wyeth. Seqslam: Visual route-based navigation for sunny summer days and stormy winter nights. In *icra*, 2012. 1, 4

[14] T. Mouats, N. Aouf, A. D. Sappa, C. Aguilera, and R. Toledo. Multispectral stereo odometry. *IEEE Transactions on Intelligent Transpotation Systems*, 2014. 4

[15] D. Nister and H. Stewenius. Scalable recognition with a vocabulary tree. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2006. 6

[16] G. Pandey, J. R. McBride, and R. M. Eustice. Ford campus vision and lidar data set. *The International Journal of Robotics Research*, 2011. 4

[17] S. Prakash, P. Y. Lee, T. Caelli, and T. Raupach. Robust thermal camera calibration and 3d mapping of object surface temperatures. In *Defense and Security Symposium*, 2006. 2

[18] M. Smith, I. Baldwin, W. Churchill, R. Paul, and P. Newman. The new college vision and laser data set. *The International Journal of Robotics Research*, 2009. 4

[19] J. Sturm, S. Magnenat, N. Engelhard, F. Pomerleau, F. Colas, W. Burgard, D. Cremers, and R. Siegwart. Towards a benchmark for rgb-d slam evaluation. In *Proceedings of the Robotics: Science and Systems Conference*. 4, 5

[20] 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. *Instrumentation and Measurement, IEEE Transactions on*, 2012. 2

[21] S. Wang, Q. Wu, X. He, and M. Xu. On splitting dataset: Boosting locally adaptive regression kernels for car localization. In *International Conference on Control Automation Robotics and Vision*, 2012. 4