2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 25-29, 2020, Las Vegas, NV, USA (Virtual)

CUHK-AHU Dataset: Promoting Practical Self-Driving Applications in the Complex Airport Logistics, Hill and Urban Environments

Wen Chen, Zhe Liu, Hongchao Zhao, Shunbo Zhou, Haoang Li, Yun-Hui Liu

***Abstract*— This paper presents a novel dataset targeting three types of challenging environments for autonomous driving, i.e., the industrial logistics environment, the undulating hill environment and the mixed complex urban environment. To the best of the author's knowledge, similar dataset has not been published in the existing public datasets, especially for the logistics environment collected in the functioning Hong Kong Air Cargo Terminal (HACT). Structural changes always suddenly appeared in the airport logistics environment due to the frequent movement of goods in and out. In the structureless and noisy hill environment, the non-flat plane movement is usual. In the mixed complex urban environment, the highly dynamic residence blocks, sloped roads and highways are included in a single collection. The presented dataset includes LiDAR, image, IMU and GPS data by repeatedly driving along several paths to capture the structural changes, the illumination changes and the different degrees of undulation of the roads. The baseline trajectories are provided which are estimated by Simultaneous Localization and Mapping (SLAM).**

I. INTRODUCTION

In the field of autonomous driving, public datasets provide good data sources for research and verification of various algorithms. They are generally collected from real environments with different characteristics using data acquisition vehicles equipped with different combinations of sensors. Among them, cameras or LiDARs as essential parts are commonly combined with other sensors to obtain information about the surroundings and the vehicles' motion. Cameras get rich texture information from the environment, and LiDARs provide 3D structure information of the environment.

Many research groups have opened up datasets with different features. The most representative and influential one is the KITTI dataset proposed in [1]. It provides LiDAR and image data in lowly dynamic rural areas and highways for different research topics, such as state estimation, 3D object detection and tracking. The dataset presented in [2] consists of two levels of sensor data collected from the complex urban environments. Focusing on the long-term autonomy, the NCLT dataset in [3] consists of omnidirectional image and LiDAR data collected along various trajectories in the campus environment. Similarly, the Oxford RobotCar Dataset [4] provides LiDAR and image data along the same route

This work was supported in part by the Natural Science Foundation of China under Grant U1613218, in part by the Hong Kong ITC under Grant ITS/448/16FP, in part by the National Key Research and Development Program of China under Grant 2018YFB1309300, in part by the Hong Kong Centre for Logistics Robotics, and in part by the VC Fund 4930745 of the CUHK T Stone Robotics Institute.

W. Chen, Z. Liu, H. Zhao, S. Zhou, H. Li and Y.-H. Liu are with the T Stone Robotics Institute and Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, HKSAR, China. (wenchen@mae.cuhk.edu.hk)

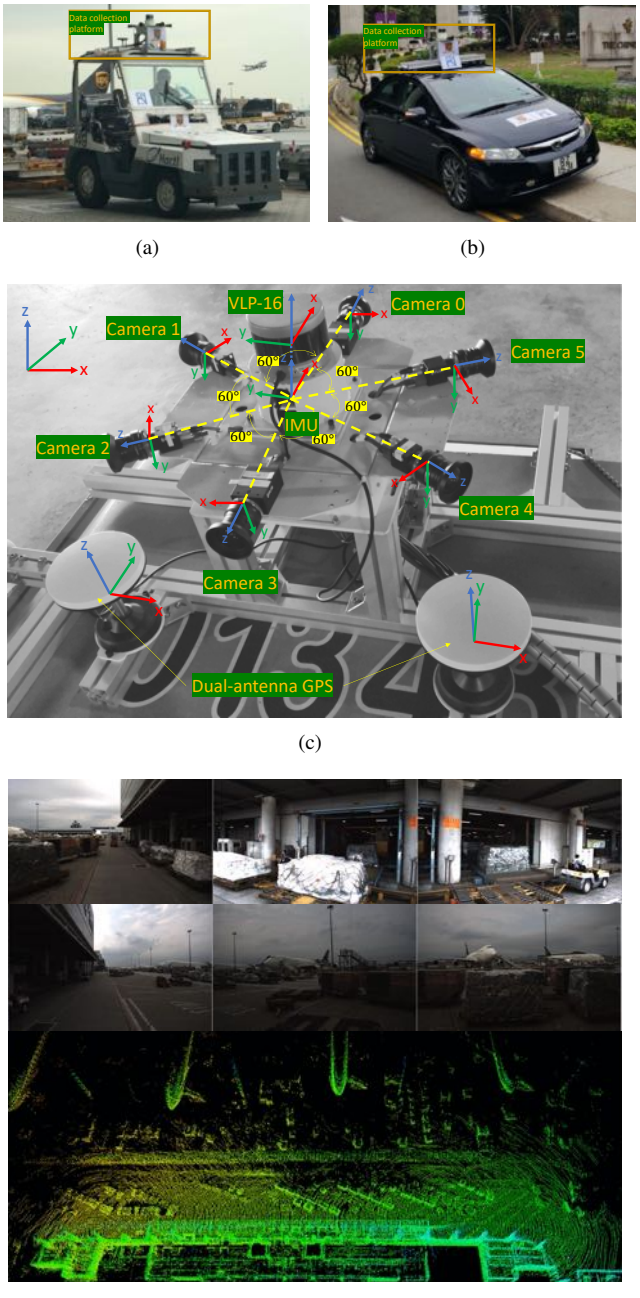


Fig. 1. Two types of data collection vehicles with the same platform. (a) Platform installed on the industrial tractor in airport. (b) Platform installed on the passenger car. (c) The layout of the sensors on data collection platform. (d) The typical scene in the industrial logistics environment. The six-camera unit provides synchronized 360° visual measurements. The 3D LiDAR (VLP-16) provides 360° range measurements.

in dynamic urban environments. Both datasets are collected at different time over one year, suffering from environment structural changes and various weather conditions. Other relevant datasets released in [5], [6], [7] are collected from urban environments using different sensor combinations.

These datasets have played a very good role in promoting the development of autonomous driving in urban or suburban environments. However, the designed collection routes are relatively simple such that few datasets can cover different environments in a single collection. What's more, to the best of our knowledge, there is not dataset that targets the challenges in the industrial logistics environment and undulating hill environment. In these two environments, autonomous driving faces different challenges compared with existing datasets. Structural changes are always taking place in the logistics environment. The environment may become completely different as goods are moved in and out at any time which makes robust localization and mapping very difficult. In the undulating hill environment, the movement of vehicles is not on a flat plane, and the surroundings are structureless and noisy.

In this paper, we present a LiDAR and image dataset that focuses on three types of environments, i.e., air cargo terminal environment, the undulating hill environment and the mixed complex urban environment using two different vehicles shown in Fig. 1. All sensors are mounted on a detachable data collection platform shown in Fig. 1(c), which is then installed on the top of the vehicles. The industrial tractor in Fig. 1(a) is used to collect data in the HACT which is the second busiest air cargo terminal in the world. The passenger car in Fig. 1(b) is used to collect data around the Chinese University of Hong Kong (CUHK) campus which is located on a hill. As for the mixed complex urban environment, we also use the passenger car to collect data along the route through the highly dynamic residence blocks, sloped roads and highways. We repeatedly collect LiDAR, camera, IMU, and GPS data along several paths in these environments to capture the structural changes, the illumination changes and the different degrees of undulation of the road. In summary, the key contributions of this paper are listed in the following:

- provides a novel dataset which firstly covers three types of challenging environments: the highly dynamic industrial logistics environment, the undulating hill environment and the mixed complex urban environment.
- provides baseline trajectories generated by the proposed SLAM approach, which combines the state-of-art LiDAR odometry, graph-based optimization and point cloud based place recognition.

The first batch of the presented dataset is available at: `http://ri.cuhk.edu.hk/research/public_datasets/CUHK_AHU_Dataset`. More data and development tools will be released periodically.

II. Collection Platform Configuration

A. Hardware Platform

Fig. 1(c) shows the sensors' locations on the data collection platform. All sensors are mounted on a carrier-independent mechanical structure which is attached on the top of vehicles. The main objective of the modular design is to acquire ominidirectional visual and range information from totally different application environments, such as airport cargo terminal and undulating hill roads. The configuration of the sensors is summarized in Table. I.

A 3D LiDAR ("VLP16") is installed at the center of the platform to provide 360-degree and 16-channel range information of the surrounding scene at a frequency of 10Hz. For the inertial measurements, a consumer-level MTi IMU sensor is mounted under the LiDAR by coaxial installation. It can provide nine-axis measurements: linear acceleration, rotational angular velocity and geomagnetism orientation. The platform also has a dual-antenna GPS with RTK correction signals to provide 2.5Hz position measurement.

The platform includes six color cameras on a flat plane: two Point Grey Grasshopper3 cameras and four Point Grey Blackfly cameras. The two Grasshopper3 cameras provide the front and rear view for better imaging quality, and the remaining four Blackfly cameras are evenly distributed in the other view directions around the central axis of the LiDAR to form an ominidirectional visual sensing unit. To perform the time synchronization of the six-camera unit, synchronized capture technology provided by the Point Grey is applied to ensure the same frame rates. One "primary" camera is used to trigger the other "secondary" cameras by linking their GPIO pins. In our application, the "Camera 2" is treated as "primary" camera, yielding 20Hz image capture rate.

All sensors are logged using an industrial computer running Ubuntu Linux with an i7 processor, 32GB DDR4 memory and two 2TB SSDs. All the sensor drivers are developed on the ROS Kinetic. And the logger are based on the rosbag package. The timestamp for each sensor measurement is created by the related driver running on the computer.

B. Sensor Calibration

1) Camera Calibration: The camera calibration includes the estimation of the intrinsic parameters of each camera and the extrinsic transformation among six cameras. The intrinsic calibration is performed separately using the *Single Camera Calibrator App* developed by MATLAB. To improve the intrinsic calibration, we remove the high-error images to make the mean reprojection error lower than 0.3 pixels.

For the extrinsic calibration among cameras, the feature-matching based method is hard to be applied because of the limited overlapping view field between any two adjacent cameras. Hence, we choose to recover the transformation matrix between any two cameras from the calibration result of the corresponding LiDAR-camera pairs.

2) Joint LiDAR-Camera Calibration: The calibration problem between LiDAR and camera is a typical Perspective-n-Point (PnP) problem. One common strategy is to transfer the 3D-to-2D problem to the 3D-to-3D problem by estimating the depth of each 2D image feature point with the help of checker board. However, in practice, the placement of the checker board affects the final calibration result since the working distance of the LiDAR ranges from $1m$ to $100m$.

TABLE I

DETAILED CONFIGURATION INFORMATION OF THE SENSORS USED ON THE PLATFORM

Sensor item	Manufacturer	Model	Hz	Description
3D LiDAR	Velodyne	VLP-16	10	16-channel and 360-degree 3D LiDAR
IMU	Xsens	MTi-300	100	Full gyro-enhanced AHRS
Dual-antenna GPS	Samsung	OEM718D	2.5	Commercial grade GPS
Main camera 0	Point Grey	GS3-U3-23S6C-C	20	1920 × 1200, 1/1.2" Sony IMX174 CMOS, global shutter, 8mm anti-shock & vibration lens (Kowa 1" LM8HC-V), 79.4○ HFov
Slave camera 3	Point Grey	GS3-U3-23S6C-C	20	1920 × 1200, 1/1.2" Sony IMX174 CMOS, global shutter, 8mm lens (Kowa 1" LM8HC), 79.4° HFoV
Slave camera 1&2&4&5	Point Grey	BFLY-U3-23S6C-C	20	1920 × 1200, 1/1.2" Sony IMX249 CMOS, global shutter, 8mm lens (Kowa 1" LM8HC), 79.4° HFoV

In order to reduce the influence of this factor, we manually selected 3D-2D point pairs from different distance zone. For each LiDAR-camera pair, we adopt the same method to find the 3D-to-2D correspondences which were initially solved by the non-iterative EPnP algorithm [8]. In addition, we also select the image points associated with the same 3D LiDAR point from the overlapping area of two adjacent images to enhance the connection between cameras. Finally, all corresponding pairs are put into a jointly non-liner optimization problem solved by a Gauss-Newton scheme. More details can be found in [9].

3) GPS/IMU-LiDAR Calibration: The fusion of GPS and IMU provides estimation of the vehicle states [10]. And also, the trajectory of the vehicle can be estimated from the matching of the edge and planar features points extracted from LiDAR measurements [11]. The calibration of GPS/IMU and LiDAR becomes a well-known hand-eye calibration problem [12]. However, the accuracy of the INS and the LiDAR odometry highly depends on the surrounding scenarios: the former needs to work in open and unsheltered environments, and the latter needs environments full of structural information. Thus, we choose an outdoor parking lot to calibrate the extrinsic parameters between the GPS/IMU and LiDAR.

III. DATASET DESCRIPTION

The datasets collected using our two types of manually driving vehicles cover different features such as industrial logistics environment, undulating hill area and mixed complex urban environment. All data are grouped by the path of collection, and each group contains at least two sets of data acquired at different time from day to night. The collection paths of each group are shown in Fig. 2. And the statistics of the size, frame counts, complexity and elevation for each path are listed in Table. II.

A. Acquisition Environments

1) Functioning Air Cargo terminal: The HACT group of datasets is collected in a normally functioning air cargo terminal with the ability of handling 3.5 million tonnes of cargo every year. It contains an indoor multi-level storage system and several outdoor box storage zones, as shown in Fig. 2(a). In order to record the most real logistics scenario, we collect the data of the whole terminal environment along forward and backward paths, without special traffic control.

Fig. 3(a) shows the dramatic changes in the terminal. During data collection, goods, trailers or cargo planes parked at every location may be removed or replaced by others. In addition, the GPS signals become inevitably unstable or unavailable as the tractor frequently switched between inside and outside areas. Thus, the GPS measurements are not provided in this group.

2) Undulating hill environment: Fig. 2(b) shows three paths of collection in the CUHK campus which is situated on an undulating hill. The elevation difference of the paths can exceed one hundred meters. During the acquisition process, the average driving speed of the car was about $30km/h$. The samples of the data group are shown in Fig. 3(b).

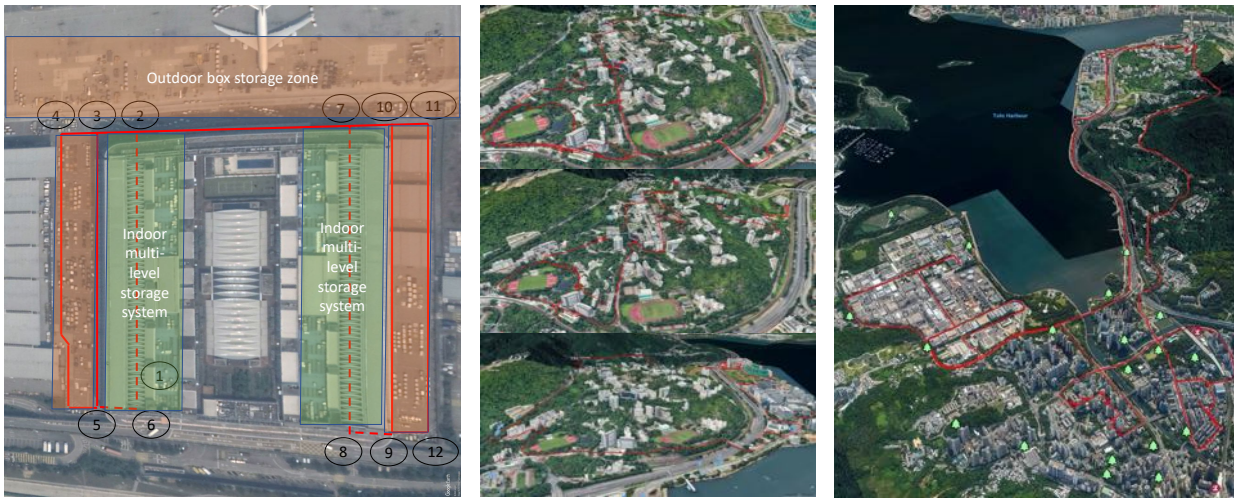
3) Mixed complex urban environment: Fig. 2(c) presents the path in mixed complex urban environments, including the non-flat mountain roads, highly dynamic living blocks, structured industrial zone and highways. The driving speed varied from $0km/h$ to $90km/h$. In the process of collection, we experienced many situations in manual driving, such as long-time traffic jam, overtaking and being overtaken. Fig. 3(c) illustrates the typical scenes in the living block and the highways.

B. Data Formats

1) Synchronized Images: All datasets are distributed by the paths of collection into groups ("HACT", "CUHK", "Taipo"). In order to reduce unnecessary trouble in the process of downloading, each set is further divided into several sub-sets whose size dose not exceeds 5GB. Such a sub-set is compressed in a tar file which is named as $\langle Group\rangle\_\langle Path\rangle\_\langle No.sets\rangle\_\langle No.subsets\rangle\ .tar$ (e.g., $HACT\_forward\_1\_01.tar$). The file tree of the sub-set is presented in Fig. 4. And the timestamps of measurements from the six cameras and velodyne LiDAR are saved respectively in the $\langle sensor\rangle\_timestamps.csv$. The formats for other measurements are as shown in the following:

All six cameras synchronize their time using a primary-secondary trigger mechanism. The timestamps of all images are also recorded in case there is a need for higher accuracy in time processing. The synchronous frequency of these cameras is $20Hz$. The images are saved in the lossless PNG format whose bayer mode is RGGB8. To convert the bayer images to RGB images, the *demosaic* function in MATLAB or the *cvtColor* function in OpenCV can be applied.

2) 3D LiDAR scans: The 3D LiDAR scans assembled from the packets returned by the VLP-16 are saved as binary files at a frequency of $10Hz$. The timestamp of the last packet during a rotation are recorded as the timestamp of the scan which is used to name the saved file: $\langle Timestamp\rangle\ .bin$. Users should pay some attention on the



(a) (b) (c)

Fig. 2. The paths of data collection. (a) The path of data collection in HACT using the industrial tractor. And the numbers are used to visualize the sequence of the forward and backward paths. The forward sequence is $1-2-3-4-5-6-2-3-5-6-1-2-7-8-9-10-7-8-9-12-11-10-7-2-1$, and the backward sequence is $1-2-7-10-11-12-9-8-7-10-9-8-7-2-1-6-5-3-2-1-6-5-4-3-2-1$. (b) The path of data collection in CUHK campus using the passenger car. (Top): CUHK Path 1; (Middle): CUHK Path 2; (Bottom): CUHK Path 3; (c) The path of data collection in Taipo using the passenger car. The paths are represented by GPS positions.

TABLE II
STATISTICS OF THE WHOLE DATASET

Group	Path	No. Sets	Size	Duration(s)	No.of images/clouds	Complexity	Elevation (Min/Avg/Max)
HACTL	Forward	8	2922.0 GB	10173	205200/101689	*****	0 meters
		8	2771.5 GB	9653	192709/96461	*****	0 meters
CUHK	Path 1	6	1558.2 GB	9671	193577/96373	***	5/37/123 meters
		6	1541.3 GB	9852	198092/98226	***	6/61/133 meters
		4	2591.1 GB	9003	181576/90010	***	4/23/128 meters
Taipo	Path 1	2	2098.7 GB	10734	214428/107295	*****	1/18/127 meters

recovering of the cut angle since it is not specified during collection. All the points are represented by four-dimension floating tuples: (X, Y, Z, I). I is the intensity of the point. (X, Y, Z) is the coordinate value of the point represented in the local Cartesian coordinate defined w.r.t moving LiDAR, at the time corresponding to the recorded timestamp. So the motion distortion problem of the 3D pointclouds is always present, and we haven't done any process to compensate them in the raw data.

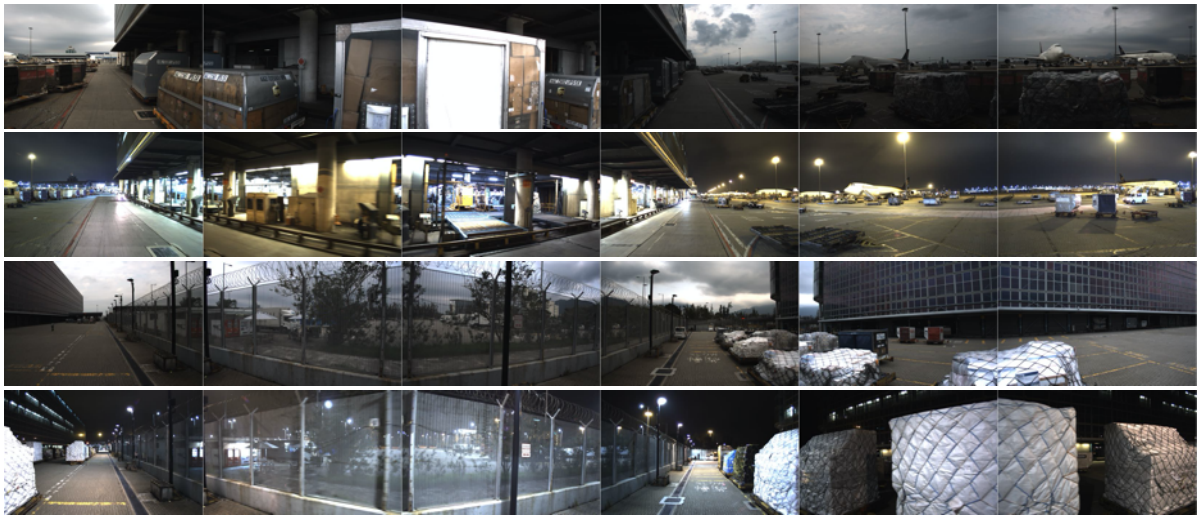
3) GPS/IMU: The GPS measurements are recorded in the $gps.csv$ at $2.5Hz$. They are formated as seven-dimension tuple: (timestamp, latitude, longitude, altitude, latitude_std, longitude_std, altitude_std). The $imu.csv$ stores the IMU measurements at $100Hz$. IMU provides the nine-axis measurements from the accelerator, gyroscope and magnetometer. They are formated as (timestamp, acc_x, acc_y, acc_z, gyr_x, gyr_y, gyr_z, ori_qua_w, ori_qua_x, ori_qua_y, ori_qua_z). We provide the two-minute data for each dataset when the vehicle is stationary before moving. The initial value for the estimation of the bias and gravity can be computed. The fusion results of GPS and IMU are recorded in the $ins.csv$.

IV. BASELINE TRAJECTORIES

The ground-truth trajectories provided by high-precision sensors are always needed for evaluating the performance of different algorithms in public datasets. However, in our acquisition environments, the GPS-based positioning methods are not applicable because they are easily affected by the surrounding buildings and trees. Some paths are even indoors. Similar to [2], we provide baseline trajectories estimated by the graph-based SLAM technology [13].

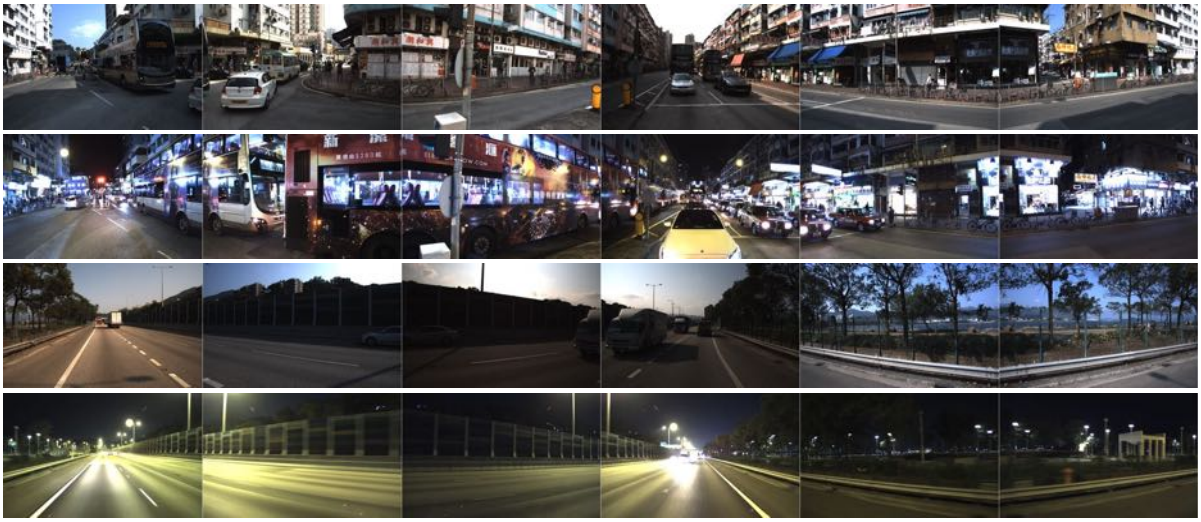
Our SLAM system fuses the sensor information from LiDAR, IMU and GPS. The measurements from IMU are initially applied to compensate the motion distortion of the LiDAR points. Similar to the state-of-art LiDAR odometry [11], the planar and edge features extracted from the undistorted pointclouds are used to establish the constrains with adjacent frames or local map clouds. The integral of IMU data are used as the initial guess of the optimization problem constructed from the constrains.

The back-end pose graph optimization is then performed to get the globally consistent map. The key to this optimization problem is not only to obtain the estimation of the relative transformations between pose nodes, but also to obtain the closed-loop constraints by place recognition. In our previous work, a deep neural network LPN-Net [14] was proposed to achieve the point cloud based place recognition in large scale environments. In this paper, we encode the point cloud of each key frame to a global descriptor using the LPN-Net. The re-passed places can be recognized by comparing the Euclidean distance between the newly-added frame and the old key frames. Then the relative transformation between the



(a)

(b)



(c)

Fig. 3. The image samples of each group. The columns from left to right represent the images captured by cameras 0 to camera 5. (a) The samples of the HACT group, presenting the dramatic changes happened every day in the air cargo terminal. (b) The samples of the CUHK group, illustrating the ups and downs of the road in the hill environment. (c) The samples of the Taipo group, showing the crowded roads and the high-speed roads in the mixed complex urban environment.

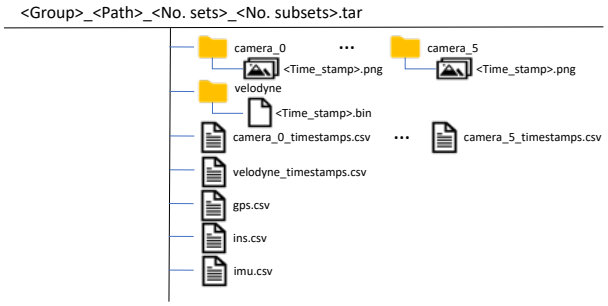
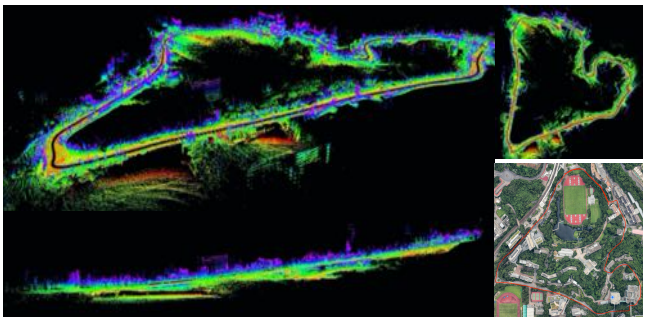


Fig. 4. File tree of each sub-set.

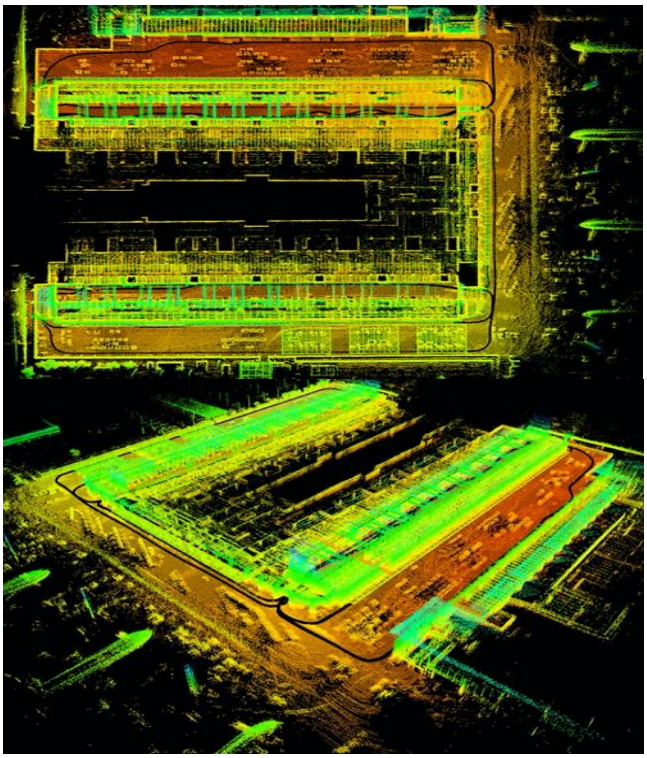
matched clouds are estimated by GICP [15] with the initial guess from the GPS measurements if they are available. The reconstructed point-cloud maps in two typical scenarios are presented in Fig. 5. And the optimized trajectories treated as baseline are presented by the black points in the map.

V. Conclusion and Future Work

We presented a new dataset focusing on autonomous driving in the industrial logistics environment, the undulating hill environment and the mixed complex urban environment. Due to the unavailability of the GPS in some collection paths, we provide baseline trajectories estimated by the fusion of LiDAR, IMU and GPS using SLAM. In the future, we will improve the accuracy of the baseline trajectories by adding image data. The extrinsic calibration between the LiDAR and six cameras will be improved with the consideration of the influence of the partially overlapping among cameras. The annoation works, like the joint objects annotation on the ominidirectional 2D visual and 3D range measurements, will be added in the dataset. In order to accelerate the application of autonomous driving technology in the industry, the dataset will continue to be updated, and more typical industrial environments will be added.



(a)



(b)

Fig. 5. The reconstructed maps of part of CUHK (a) and HACT (b). The baseline trajectories are represented by black points in the maps.

References

[1] 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, pp. 3354–3361.

[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," *International Journal of Robotics Research*, vol. 38, no. 6, pp. 642–657, 2019.

[3] 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, 2016.

[4] W. Maddern, G. Pascoe, C. Linegar, and P. Newman, "1 year, 1000 km: The Oxford RobotCar dataset," *International Journal of Robotics Research*, vol. 36, no. 1, pp. 3–15, 2017.

[5] X. Huang, X. Cheng, Q. Geng, B. Cao, D. Zhou, P. Wang, Y. Lin, and R. Yang, "The apolloscape dataset for autonomous driving," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 1067–1073.

[6] A. S. Huang, M. Antone, E. Olson, L. Fletcher, D. Moore, S. Teller, and J. Leonard, "A high-rate, heterogeneous data set from the DARPA urban challenge," *International Journal of Robotics Research*, vol. 29, no. 13, pp. 1595–1601, 2010.

[7] M. Smith, I. Baldwin, W. Churchill, R. Paul, and P. Newman, "The new college vision and laser data set," *International Journal of Robotics Research*, vol. 28, no. 5, pp. 595–599, 2009.

[8] V. Lepetit, F. Moreno-Noguer, and P. Fua, "Epnp: An accurate o (n) solution to the pnp problem," *International Journal of Computer Vision*, vol. 81, no. 2, p. 155, 2009.

[9] Y. Liu, C. Suo, Z. Liu, and Y. H. Liu, "A multi-sensor fusion based 2D-Driven 3D object detection approach for large scene applications," in *Proceedings of the IEEE International Conference on Robotics and Biomimetics*, 2019, pp. 2181–2188.

[10] T. Moore and D. Stouch, "A generalized extended Kalman filter implementation for the robot operating system," *Advances in Intelligent Systems and Computing*, vol. 302, pp. 335–348, 2016.

[11] J. Zhang and S. Singh, "LOAM: Lidar Odometry and Mapping in Real-time," in *Proceedings of Robotics:Science and Systems*, 2014, pp. 9–17.

[12] F. Furrer, M. Fehr, T. Novkovic, H. Sommer, I. Gilitschenski, and R. Siegwart, *Evaluation of combined time-offset estimation and hand-eye calibration on robotic datasets.* Cham: Springer International Publishing, 2017.

[13] G. Grisetti, R. Kummerle, C. Stachniss, and W. Burgard, "A tutorial on graph-based SLAM," *IEEE Intelligent Transportation Systems Magazine*, vol. 2, no. 4, pp. 31–43, 2010.

[14] Z. Liu, S. Zhou, C. Suo, P. Yin, W. Chen, H. Wang, and Y.-H. Liu, "LPD-Net: 3D point cloud learning for large-scale place recognition and environment analysis," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 2831–2840.

[15] A. V. Segal, D. Haehnel, and S. Thrun, "Generalized-ICP (probailistic ICP tutorial)," in *Proceedings of Robotics: science and systems*, 2009.