PAIR360: A Paired Dataset of High-Resolution 360° Panoramic Images and LiDAR Scans

Geunu Kim, Daeho Kim, Jaeyun Jang, and Hyoseok Hwang

***Abstract*—The 360° camera is a compact omnidirectional perception system for capturing panoramic images with the same field of view as LiDAR. This boosts its versatility for use in autonomous driving and robotics. However, most existing datasets of 360° panoramic images primarily focus on indoor or virtual environments, or they offer only low-resolution outdoor images and LiDAR configurations. In this letter, we present PAIR360, a multi-modal dataset encompassing high-resolution 360° camera images and 3D LiDAR scans, aimed at stimulating research in computer vision. To this end, we collected a comprehensive dataset at Kyung Hee University Global Campus, capturing 52 sequences from 7 different areas under diverse atmospheric conditions, including sunny, cloudy, and sunrise. The dataset features 8K resolution panoramic imagery, six fisheye images, point clouds, GPS, and IMU data, all synchronized using LiDAR timestamps and calibrated across visual sensors. We also provide additional data, such as depth maps, segmentation, and 3D maps, to demonstrate the feasibility of our dataset and its application to various computer vision tasks. The dataset is available for download at: https://airlabkhu.github.io/PAIR-360-Dataset/**

***Index Terms*—Data Sets for SLAM, Sensor Fusion, Omnidirectional Vision, Data Sets for Robotic Vision**

I. INTRODUCTION

An omnidirectional perception system is crucial for recognizing and interacting with the environment in various fields, including robotics [1], autonomous driving [2], and augmented/virtual reality [3]. These systems rely on sensors such as multi-cameras, 360° cameras, and LiDAR. The adoption of 360° cameras, including fisheye cameras, is growing in these areas due to their compact and cost-effective design [4]. The typical representation of the 360° visual perceptions is in the form of 360° panoramic images acquired through equirectangular projection from multiple cameras. This image, with its omnidirectional field of view (FoV), captures more features and objects continuously, regardless of the camera's direction, offering significant advantages. It also results in more consistent and robust feature tracking [5], [6] as it remains stable even when objects are too close to the camera or lack sufficient textures by utilizing textures from other directions.

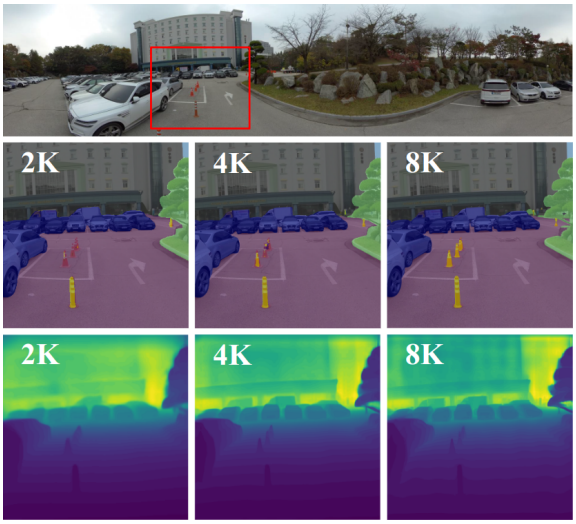


Fig. 1: Performance of segmentation and depth estimation at various resolutions. Top: panoramic input. Middle: segmentation results from panoramic input. Bottom: depth map from cropped panoramic input.

Despite these advantages, the availability of 360° panoramic image datasets for autonomous driving is limited. Commonly used datasets contained 360° panoramic images [7]–[10] are limited in their applicability to outdoor autonomous driving because they primarily focus on indoor image processing and do not effectively capture dynamic objects.

While datasets such as Ford Campus dataset [11] and PanoVILD [12] offer 360° panoramic images and LiDAR scans in outdoor settings, the resolution of these 360° panoramic images is limited. Low-resolution images lack high-frequency details, making it challenging to accurately interpret complex scenes, such as obstacles or pedestrians, which are crucial for autonomous driving [13]. This limitation is particularly critical in wide FoV conditions, such as 360° panoramic images, where higher resolution is necessary to maintain detailed information about objects [14], [15]. Fig. 1 illustrates the performance of segmentation and depth estimation for 360° panoramic images at varying resolutions, utilizing Depth-Anything [16]. The figure shows that as resolution increases, depth of field detail improves and smaller objects can be better distinguished.

The visual quality of image stitching is as critical as the res-

Manuscript received: June, 3, 2024; Revised August, 14, 2024; Accepted September, 3, 2024.

This paper was recommended for publication by Editor Pascal Vasseur upon evaluation of the Associate Editor and Reviewers' comments.

This work was partly supported by a National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (NRF-2022R1C1C1008074), and by an Institute of Information and Communications Technology Planning and Evaluation (IITP) grant funded by the Korean government (MSIT) (No.RS-2022-00155911, Artificial Intelligence Convergence Innovation Human Resources Development (Kyung Hee University), and by the Convergence security core talent training business support program(IITP-2023-RS-2023-00266615). (Corresponding author: Hyoseok Hwang.)

The authors are with the Department of Software Convergence, Kyung Hee University, Yongin-si, Gyeonggi-do, 17104, Republic of Korea (e-mail: kyak17232@khu.ac.kr; kdh2769@khu.ac.kr; yoon2926@khu.ac.kr; hyoseok@khu.ac.kr)

Digital Object Identifier (DOI): see top of this page.

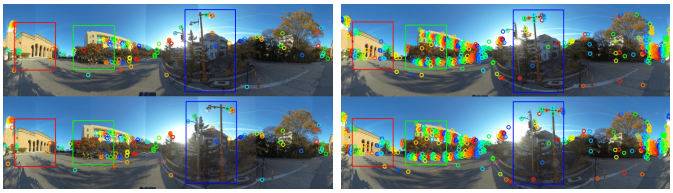
olution of the 360° panoramic image. Since the stitched image is composed of images from multiple cameras, variations in brightness and color balance can occur due to differences in exposure, white balance, and other factors. Additionally, stitching algorithms that rely solely on photometric features can introduce disjoint overlapping regions between two images, known as seams. These artifacts cause repeated appearances and misalignment of objects at stitching boundaries, which can negatively affect feature matching as shown in Fig. 2.

High-resolution and high-quality panoramic images are essential for autonomous driving, but there is room for performance improvement and reliability. For example, in the case of simultaneous localization and mapping (SLAM) tasks, the performance can be enhanced by utilizing sensor data such as LiDAR, GPS, or IMU [17]. To fully leverage these advantages, sensor data is becoming an increasingly significant part of datasets. Moreover, effective use of these data requires proper sensor calibration and time synchronization [18].

Furthermore, the demand for diverse annotations in vision tasks is steadily increasing as they are crucial for improving model performance. Recent image-based studies [16], [19] propose methods that leverage annotation data, such as using segmentation labels, to improve the performance of depth estimation. Consequently, datasets providing annotation data such as depth maps or segmentation labels are increasing [7], [20]. However, there are very few outdoor panoramic image datasets that offer such diverse annotation data [11], [12], [21].

To address these issues, we present a paired dataset using a 360° camera and LiDAR system. The panoramic images we provide are 8K high-resolution and stitched without disjointed overlaps. The high-resolution and high-quality 360° panoramic images are also calibrated and synchronized with other devices such as LiDAR, GPS, and IMU. To enhance our dataset's diversity, we conducted multiple drives around Kyung Hee University Global Campus under varying atmospheric conditions, collecting approximately 12 TB of data as shown in Fig. 3. Additionally, we leverage a largely pre-trained foundation model [16] to generate dense depth maps and segmentation labels for image annotations. Finally, we assess the dataset using SLAM and multi-view stereo (MVS) techniques to verify the utility of image-paired LiDAR data within the dataset. Our main contributions can be summarized as:

- We provide seamless 8K (7680×3840) equirectangular images to preserve detailed information and support high-resolution image processing research. These 360° panoramic images are synchronized with LiDAR and share the same FoV.
- To accommodate varying weather and conditions, our dataset was collected multiple times at Kyung Hee University Global Campus, covering various atmospheric scenarios such as sunny, cloudy, and sunrise.
- Outdoor panoramic datasets provide LiDAR, GPS, and IMU as well as depth maps, segmentation labels and point cloud data, which can be useful for advanced vision tasks that require outdoor environmental understanding and learning.



(a) (b)

Fig. 2: Comparison of feature matching results between fragmented and seamless panoramic images. (a) Panoramic image with visible stitching fragments [12]. (b) Our seamless panoramic image.

II. RELATED WORK

This section examines datasets relevant to 360° panoramic images. Table I provides a summary of significant datasets containing 360° panoramic images and autonomous driving dataset.

A. 360° Image Dataset

Matterport3D [7] and 2D-3D-S dataset [9] are widely used across various applications because they include images and diverse metadata such as normal images, depth images, and semantic segmentation. However, these datasets are focused on indoor environments, and their application to outdoor environments is challenging due to domain gaps.

There are few equirectangular outdoor datasets available because of the limitations of sensors. 360VO [25], and 360VIO [5] datasets are equirectangular datasets for visual odometry. Despite offering continuous scenes and IMU data, these outdoor datasets do not provide depth information. The Depth360 dataset [21] is utilized for estimating depth from monocular 360° videos. It generates depth maps through a learning-based approach under certain constraints, providing outdoor equirectangular images and ground truth depth maps. The dataset comprises four pairs of images captured by rotating the camera 90° at the same locations. However, it is unsuitable for 3D reconstruction and visual odometry because the images were taken independently at multiple locations. Additionally, the depth maps lack fine-grained representations due to their indirect generation method.

B. Autonomous Driving Dataset

A widely utilized real-world dataset is KITTI-360 [20], which features a 180° FoV fisheye camera on each side. Other notable datasets incorporating multiple cameras include nuScenes [26], Oxford Robot Car [27], and WoodScape [28]. These datasets have significantly contributed to the cognitive capabilities of autonomous vehicles by providing valuable resources such as segmentation labels, bounding boxes, and vehicle bus signals. However, utilizing more cameras is not always ideal because it increases cost and complexity in practical applications. This approach necessitates extra effort for camera calibration to determine the relative positions between images, followed by image-to-image feature matching and filtering.

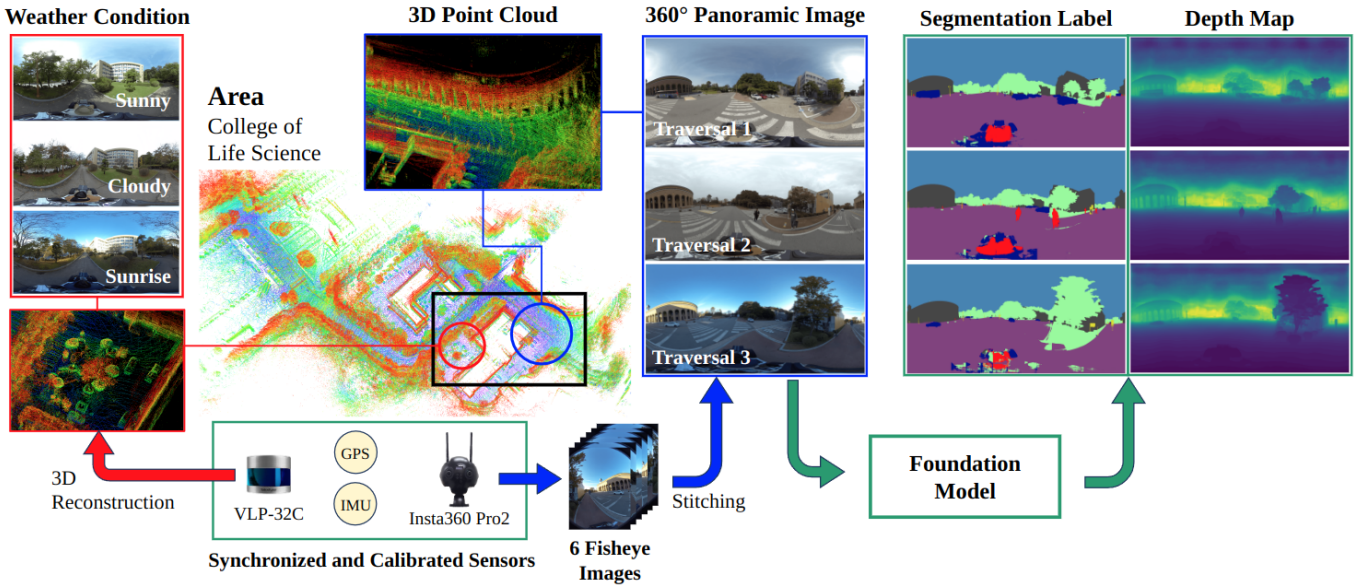


Fig. 3: Overview of our dataset. This figure shows *College of Life Science* area by equirectangular images, point cloud map, segmentation labels and depth maps. Two equirectangular image pairs show each location for three different traversals and the weather at the time. The point cloud map is generated by FAST-LIO-SLAM [22]. Annotation data include segmentation labels and depth maps are generated by Depth-Anything [16].

TABLE I: Summary of 360° image datasets including the availability of panorama images, the resolution of 360° panoramic or fisheye images, the availability of LiDAR sensing data, depth maps, semantic segmentation labels, spatial continuity of frames, the number of 360° panoramic images, and the total number of RGB images (indicated in brackets).

Dataset Name	Setting	Panorama	Resolution	LiDAR	Depth	Segmentation	Continuity	# Images
KITTI-360 [20]	Outdoor	✗	1400×1400	✓	✓	✓	✓	0 (332,000)
AMUSE [23]	Outdoor	✗	616×1616	✓	✗	✗	✓	0 (704,640)
360VOT [24]	Both	✓	3840×1920	✗	✗	✗	✓	113,000 (113,000)
Matterport3D [7], [8]	Indoor	✓	2048×1024	✗	✓	✓	✗	9,658 (209,058)
2D-3D-S [9]	Indoor	✓	4096×2048	✗	✓	✓	✗	1,413 (71,907)
Depth360 [21]	Outdoor	✓	1440×720	✗	✓	✗	✗	30,000 (30,000)
Ford Campus [11]	Outdoor	✓	1616×3080	✓	✗	✗	✓	7,789 (38,945)
PanoVILD [12]	Outdoor	✓	2000×1000	✓	✗	✗	✓	15,443 (92,658)
PAIR360 (ours)	Outdoor	✓	7680×3840	✓	✓	✓	✓	88,222 (617,554)

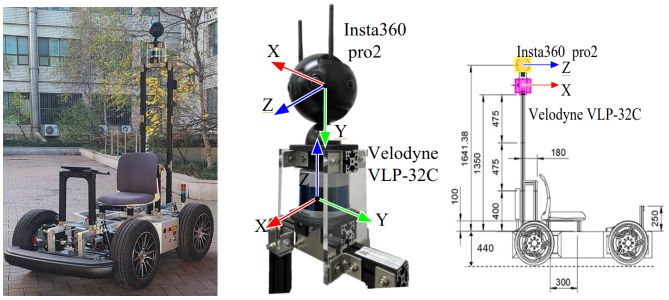
Conversely, several datasets [11], [12], [23], [29] employ a 360° camera. The AMUSE dataset and NCLT dataset [23], [29] employ the Ladybug3 360° camera, GPS, and IMU to deliver synchronized driving data under various weather conditions. However, these datasets only provide fisheye images without equirectangular images, necessitating image stitching as a preliminary step.

The Ford Campus dataset [11] employed a Ford truck outfitted with 360° cameras, 3D LiDAR, and forward-looking LiDAR for SLAM, three-dimensional perception, and object detection. This dataset was collected by driving around the campus and downtown areas, providing both undistorted images from each lens of the 360° camera and a composite 'full image' of these lenses stacked together. The PanoVILD dataset [12] utilizes a 360° camera, 3D LiDAR, IMU, and RTK GPS, offering data that includes multiple small, large, and off-road sequences with loop closures for visual SLAM and visual-LiDAR odometry. However, these datasets have a lower resolution than those with a wide FoV and produce noise due to luminance differences around the stitching boundaries. In particular, the Ford Campus datasets can suffer from stacking issues without accounting for parallax between images, resulting in discontinuous appearances of objects. Moreover, these datasets lack multiple iterations of the same scene, which is necessary to evaluate the robustness of the models under varying weather conditions.

III. HARDWARE SETUP AND CALIBRATION

A. Sensor Setup

The 360° camera (Insta360 Pro2) and the 3D mechanical LiDAR (Velodyne VLP-32C) were mounted vertically on the mobile platform ERP-40. The sensors were positioned approximately 190 cm above the ground, which is the height of a typical car. The hardware setup is shown in Fig. 4.



(a) (b) (c)

Fig. 4: Hardware setup for data collection. (a) The mobile platform. (b) The sensor setup. (c) Sensor location diagram.

TABLE II: Overview of the sensors' specifications.

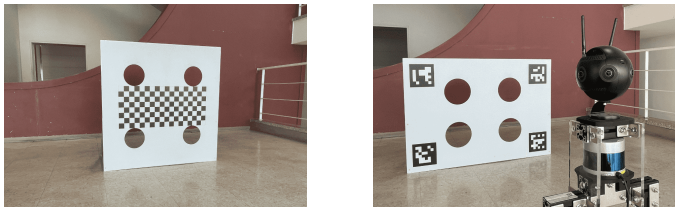
Sensor	Characteristics	Rate
Camera	200° F2.4 fisheye lens×6, auto exposure, 1/2.3" CMOS sensor, rolling shutter	30 Hz
LiDAR	32 channels, 200 m measurement range, 360° horizontal and 40° vertical FoV, collecting 1.2 million points/second, ±3 cm range accuracy	10 Hz
GPS	2-2.5 m horizontal positioning accuracy, 0.05 m/s velocity accuracy, 0.3° heading accuracy	10 Hz
IMU	9-axis	500 Hz

Additionally, we utilized the GPS and IMU sensors built into the Insta360 Pro2 camera to construct the dataset. Using the open-source robot operating system (ROS) [30], we collect LiDAR scans and print the timestamps of LiDAR data for synchronization with the camera. The detailed sensor specifications are described in Table II.

B. Sensor Calibration

1) Fisheye Camera Intrinsic Calibration: We employ an equidistant distortion model [31] for the intrinsic calibration of fisheye lenses. Using a 7×17 array of checkerboards in Fig. 5a with a grid size of 50 mm, we determine two sets of parameters for each camera lens: intrinsic matrix K and distortion coefficients D.

2) Sensors Extrinsic Calibration: Our extrinsic calibration approach estimates the rigid transformations between the six fisheye cameras, the LiDAR, and the built-in IMU. We use the first camera on the right side at the front of the multi-camera system as a reference to estimate the relationships among the other sensors. First, we estimate the transformation matrix between the reference camera and the LiDAR. For extrinsic calibration of the LiDAR and a camera, we employed the method by Beltrán et al. [32] using four ArUco markers [33] and circular holes in Fig. 5b. This calibration confirmed that the LiDAR coordinate system is offset by 0.21 m along the y-axis relative to the camera coordinate system. Secondly, for the extrinsic calibration among the cameras, we utilized ArUco markers following the same methodology. Finally, we calibrated the extrinsic parameters of the IMU using the Kalibr [34]. We confirmed that the built-in IMU is positioned at the center in front of the two cameras.



(a) (b)

Fig. 5: Two types of target board for sensor calibration. (a) Setup for calibration of fisheye cameras' intrinsic parameters. (b) Setup for estimating extrinsic parameters between cameras and between camera and LiDAR.

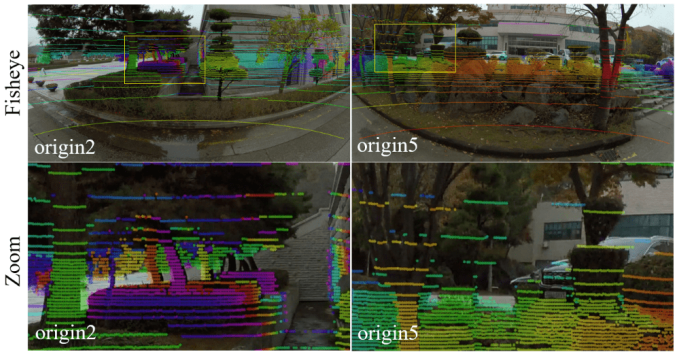


Fig. 6: The projected LiDAR point cloud onto each fisheye images (top) and partial zoomed-in images (bottom). Projected points are color-coded based on distance.

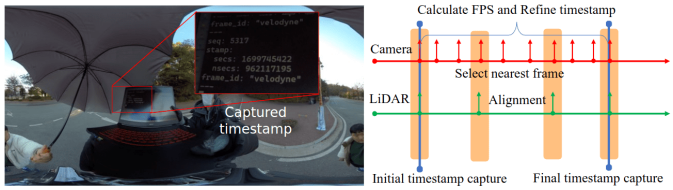
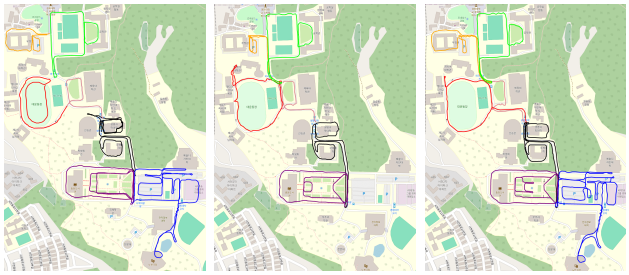


Fig. 7: Example of the synchronization method by using timestamp and frame rate.

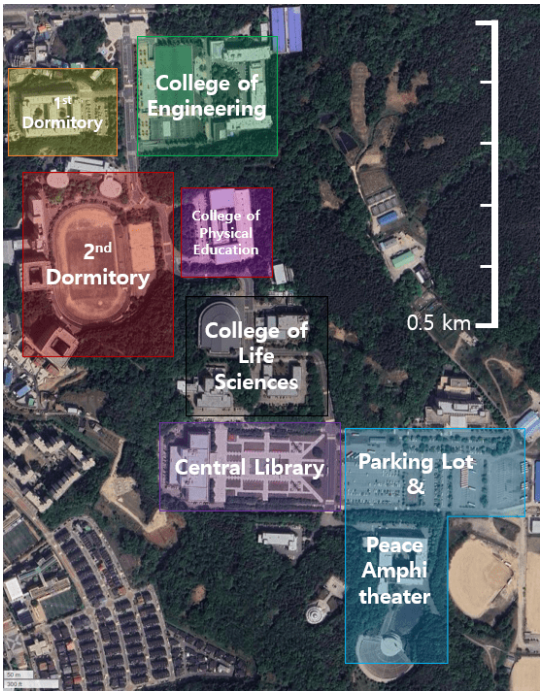
The quality of the calibration is illustrated in Fig. 6, showing the alignment of LiDAR points with the image. For clarity, only points within 30 m are projected, and the fisheye image is cropped to display the area around the projection.

C. Time Synchronization

Precise synchronization of timestamps across sensors is a critical issue when building large multi-modal datasets. In our case, since the camera lacks hardware-based synchronization, we employed an alternative method. The timestamp of the LiDAR sensor can be obtained from the ROS timer, which uses the operating system (OS) timer identically. Conversely, the 360° camera synchronizes with the operating system's time only while connected to the laptop; afterward, it uses a built-in timer for timestamps. However, we found discrepancies between the timestamps from the built-in timer and the OS's timer, prompting us to use post-processing to synchronize the multi-modal datasets.



(a) (b) (c)



(d)

Fig. 8: Visualization of the vehicle travel path by area of each traversal. (a) Traversal 1 was conducted on clear days in the afternoon, spanning multiple occasions. (b) Traversal 2 was conducted throughout the day, from morning to afternoon, on cloudy days. (c) Traversal 3 was carried out at sunrise when the sun was near the horizon. (d) All areas in a traversal.

For post-processing of synchronization, we display the LiDAR's timestamp on a screen, which the camera captures at the start and end of the operation to synchronize the LiDAR timestamps with the corresponding camera frames. This method offers two key advantages for timestamp synchronization. Firstly, by using the timestamps at the beginning and end, we can estimate the actual frames per second (FPS) of the 360° camera, which varies due to temperature fluctuations or processing delays. We then adjust the camera timestamps based on the calculated FPS using the time duration and the number of captured frames. Secondly, aligning the camera's timestamps with those of the LiDAR aids in synchronizing data that has different FPS rates. The camera and LiDAR operate at 30 FPS and 10 FPS, respectively, leading to slightly different data acquisition timings. We synchronize the datasets by selecting camera frames that have the smallest time difference from the LiDAR timestamps. Since the GPS and IMU timestamps align with those of the camera, the adjusted camera timestamps are also applied to these sensors. An illustration of how LiDAR timestamps are displayed and synchronized in multi-modal datasets can be seen in Fig. 7.

TABLE III: Dataset summary. The length and duration are calculated by summation of all sequences of each area.

Area	Traversal	Date	Condition	Length	Duration
College of Life Science	1	08-31	Cloudy	1.87 km	802 s
	2	11-04	Cloudy	1.51 km	572 s
	3	11-08	Sunrise	1.72 km	592 s
College of Physical Education	1	08-25	Sunny	0.50 km	219 s
	2	11-04	Cloudy	0.59 km	216 s
	3	11-08	Sunrise	0.52 km	216 s
Parking Lot & Peace Amphitheater	1	10-07	Cloudy	2.62 km	922 s
	3	11-07	Sunrise	2.84 km	866 s
Central Library	1	10-07	Cloudy	1.84 km	646 s
	2	11-04	Cloudy	1.74 km	521 s
	3	11-07	Sunrise	1.88 km	548 s
1st Dormitory	1	09-24	Sunny	0.64 km	224 s
	2	11-04	Cloudy	0.65 km	213 s
	3	11-12	Sunrise	1.08 km	294 s
2nd Dormitory	1	08-25	Sunny	0.97 km	282 s
	2	11-04	Cloudy	0.94 km	328 s
	3	11-08	Sunrise	0.95 km	284 s
College of Engineering	1	09-24	Sunny	1.13 km	410 s
	2	11-04	Cloudy	1.13 km	329 s
	3	11-12	Sunrise	1.09 km	328 s

IV. DATASET

A. Dataset Description

The PAIR360 dataset, comprising pairs of equirectangular images and LiDAR scans, was collected along the roads of Kyung Hee University Global Campus. We divided the campus into seven areas, with each area surveyed independently. The vehicle was manually steered using a wireless controller. Driving through all areas of the campus is called a *traversal*. To ensure the robustness and diversity of our dataset, we conducted three traversals of the campus under different conditions. It resulted in a total of over 600,000 images and 85,000 laser scans. Fig. 8 shows the data collection traversals and visualizes the area within the university using GPS signals.

The first traversal was carried out during daylight hours on August 25 and 31, September 24, and October 7, 2023, capturing the campus under clear and cloudy skies in the afternoon. The total distance covered in the first traversal, as recorded by the GPS sensor, is approximately 9.56 km, and the total time spent traveling, calculated from the video times, is 3,505 s. The second traversal was performed on November 4, 2023, under cloudy conditions throughout the day to maintain weather consistency. This traversal provides valuable insight into different illumination conditions by comparison with the other traversals. The total distance covered in the second traversal is approximately 6.56 km, with a total driving time of 2,179 s. The final traversal took place on November 7, 8, and 12, 2023, to capture the campus scene with the sun near the horizon, approximately one hour after sunrise. The total distance for the third traversal is approximately 10.09 km, and the total driving time is 3,128 s. Table III summarizes the

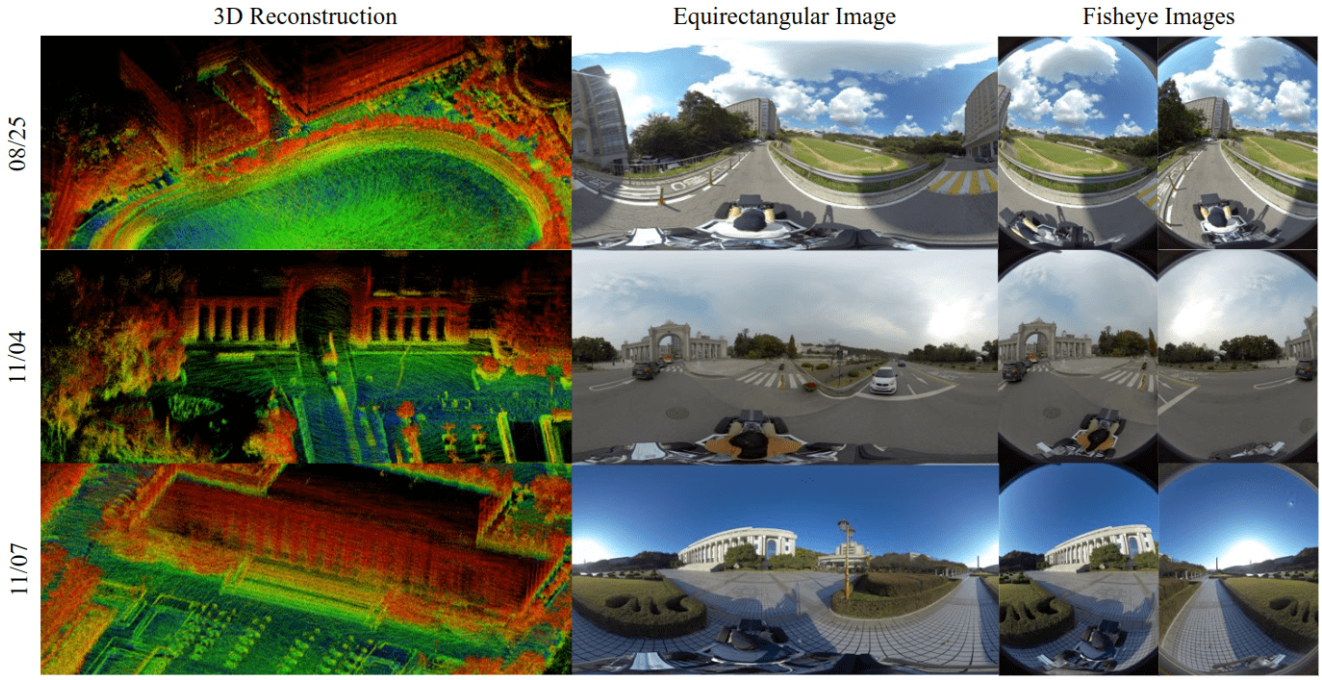


Fig. 9: Examples of 3D point cloud map, equirectangular image, and fisheye images. The map was reconstructed by FAST-LIO-SLAM.

date, weather, length, and duration of each traversal by area. For length and duration, we present the sum of the length and duration of all sequences. Fig. 9 presents sample data from the dataset collection date. The sample data includes 3D point clouds, 360° panoramic images, and fisheye images generated using LiDAR.

B. Data Description

To create a dynamic dataset, we divided the data from each area into sequences based on the vehicle's stops. Each sequence includes LiDAR scans, fisheye images, 360° panoramic images, segmentation labels, depth maps, GPS data, and IMU data. Specifically, the fisheye images are organized into folders labeled *origin1* to *origin6*, corresponding to the images captured by each of the six lenses.

1) LiDAR Data: For efficiency, we extracted Velodyne ROS bag files into point cloud scans in PCD format, which can be parsed easily using the provided Python code. Each PCD file is approximately 1.1 MB and stores the location of each point along with other fields such as intensity and ring. The PCD file name represents the timestamp that Velodyne generated the point cloud scan, recorded as UNIX time in nanoseconds since 00:00:00 January 1, 1970, Coordinated Universal Time, represented by a 19-digit integer.

2) Images: We configured the Insta360 Pro2 camera to record in '8K/30F/3D' mode, capturing 3840×2880 resolution video at 30 FPS from each of the six lenses. We extracted RGB fisheye images from these videos and saved them as PNG files of the same resolution. Additionally, we used the Insta360 Pro Stitcher program, the manufacturer's official stitching software, to create 3D stitched videos. The program was set to 'monoscopic' content type and 'new optical flow' stitching mode. The stitched images were extracted from the resulting 7680×3840 resolution video at 29.97 FPS and saved as PNG files of the same size. We named the file names as 4-digit integers starting from 0000. We also provide a "CAM.csv", which contains information on the image name, camera UNIX timestamp, and name of LiDAR scan in each row. The camera UNIX timestamp is a 19-digit integer to align with the LiDAR timestamp.

3) GPS/IMU Data: The Insta360 Pro2 camera has a built-in GPS and an IMU. GPS and IMU data are embedded in the video files. After recording, we extract the GPS and IMU data into plain text files. These data are synchronized with the camera frame and saved in the "GPS.csv" and "IMU.csv" files. Since the GPS requires signals from three satellites to estimate its current location, the "GPS.csv" file includes three timestamps, along with latitude, longitude, and altitude for each frame. The "IMU.csv" file contains synchronized timestamps, 3-axis gyroscope values, and 3-axis accelerometer values.

V. DATASET EVALUATION AND ANNOTATION

In this section, we present the results of using state-of-the-art models to evaluate and enhance our generated datasets. We compared the collected GPS data with trajectories generated by SLAM and MVS models. Additionally, we created dense depth maps and segmentation labels through deep learning network to supplement the sparsity of LiDAR data and to provide semantic information. Fig. 11 illustrates the projection of LiDAR points onto a panoramic image, along with the results of the segmentation labels and the depth map. The generated depth maps are evaluated by HoHoNet [35] and LiDAR scan.

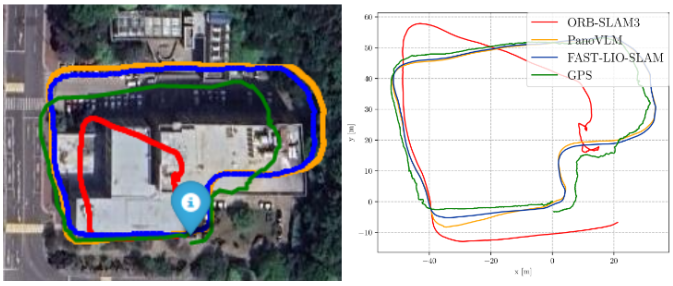


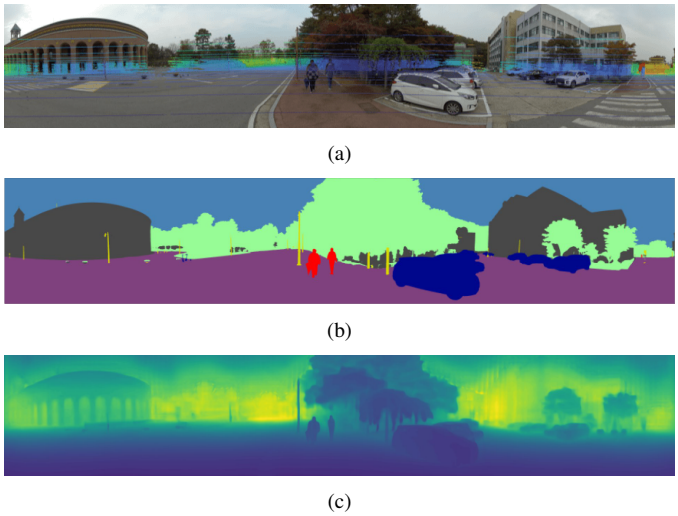
Fig. 10: Comparison trajectories of ORB-SLAM3 (red), PanoVLM (orange), FAST-LIO-SLAM (blue), and GPS (green). Trajectories are generated using *the College of Life Science* area in traversal 1.

TABLE IV: ATE translation with GPS in meters of output trajectory using *the College of Life Science* in traversal 1.

Sequence	Method	Data Type		ATE (m)	
		Image	LiDAR	RMSE	SD.
Sequence 3	ORB-SLAM3	✓		15.23	6.71
	FAST-LIO-SLAM		✓	4.82	1.92
	PanoVLM	✓	✓	5.05	1.91

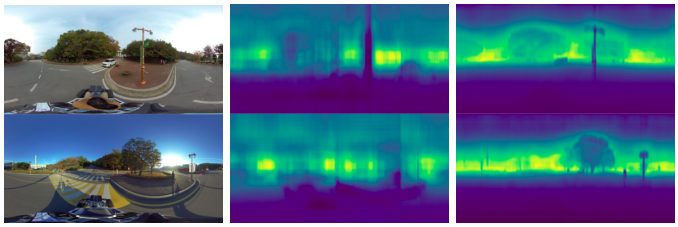
1) Localization and Mapping: The dataset was evaluated using state-of-the-art algorithms [17], [22], [36], with GPS measurements serving as the ground truth for calculating absolute trajectory error (ATE) [37]. ORB-SLAM3 [36] was conducted using stereo-fisheye images and IMU data, while FAST-LIO-SLAM [22], a feature-based SLAM system, utilized LiDAR scans and IMU data. For ORB-SLAM3, fisheye images were resized to 1280×960 pixels, and 50,000 features were extracted. In PanoVLM [17], LiDAR was employed for panoramic vision and fusion mapping, with 16,192 features extracted from a 2K 360° panoramic image. The LiDAR and camera were integrated into the SLAM system at 10 FPS, while IMU data was integrated at 500 FPS. The resulting trajectories are illustrated in Fig. 10, and the quantitative results are presented in Table IV. Sensor positions estimated to exceed 10,000 m were classified as outliers and excluded from the evaluation. ORB-SLAM3 struggles to extract features during rotational motion, which can result in errors in rotation estimation and failures in loop closure. These results underscore the importance of auxiliary data, such as LiDAR, for effective SLAM in outdoor environments. Moreover, GPS signals can be unreliable near tall buildings; therefore, we provided pose estimates along with point cloud data obtained using FAST-LIO-SLAM.

2) Segmentation Label and Depth Map: Recent research [38], [39] has demonstrated the effectiveness of training deep neural networks using foundation models with strong zero-shot capabilities and high performance. Inspired by these works, we generated segmentation labels and dense depth maps using Depth-Anything [16], which can be used as ground truths. While 8K segmentation labels were produced using the fine-tuned model, 8K depth maps were generated by interpolating outputs from 2K resolution images due to hardware constraints.



(a) (b) (c)

Fig. 11: Paired dataset with annotations. (a) The projection of the LiDAR points onto the panoramic image. (b) The segmentation label and (c) Depth maps generated by Depth-Anything [16].



(a) (b) (c)

Fig. 12: Depth estimation using HoHoNet. (a) Input images. (b) Pre-trained model. (c) Re-trained model.

We conducted further experiments to demonstrate the applicability of the additional datasets we provide. We employed HoHoNet [35], a depth estimation model with pre-trained weights from an indoor dataset. Initially, we estimated depth by applying our outdoor panoramic image to the pre-trained model. We then re-train the model using depth data generated by Depth-Anything as ground truth. As shown in Fig. 12, the results clearly show improved performance after re-training, highlighting how the additional data we provide can be helpful in enhancing vision task performance.

VI. CONCLUSIONS

In this letter, we introduce PAIR360, a multi-modal dataset containing of pairs of 360° panoramic images and LiDAR scans covering a complete 360° area. The objective is to facilitate sensor fusion of LiDAR and 360° cameras for various computer vision tasks, such as 3D mapping, outdoor depth estimation, and training and evaluation of deep neural networks. Future enhancements may involve refining segmentation labels and depth maps, as well as incorporating bounding box annotations to expand the utility of PAIR360.

REFERENCES

[1] I. Marković, F. Chaumette, and I. Petrović, "Moving object detection, tracking and following using an omnidirectional camera on a mobile robot," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, 2014, pp. 5630–5635.

[2] J. Beltrán, C. Guindel, I. Cortés, A. Barrera, A. Astudillo, J. Urdiales, M. Álvarez, F. Bekka, V. Milanés, and F. García, "Towards autonomous driving: a multi-modal 360 perception proposal," in *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2020, pp. 1–6.

[3] A. Zirakchi, C. L. Lundberg, and H. E. Sevil, "Omni directional moving object detection and tracking with virtual reality feedback," in *Dynamic Systems and Control Conference*, vol. 58288. American Society of Mechanical Engineers, 2017, p. V002T21A012.

[4] U. Kart, J.-K. Kämäräinen, L. Fan, and M. Gabbouj, "Evaluation of visual object trackers on equirectangular panorama." in *VISIGRAPP (5: VISAPP)*, 2018, pp. 25–32.

[5] Q. Wu, X. Xu, X. Chen, L. Pei, C. Long, J. Deng, G. Liu, S. Yang, S. Wen, and W. Yu, "360-vio: A robust visual–inertial odometry using a 360° camera," *IEEE Transactions on Industrial Electronics*, 2023.

[6] S. Ji, Z. Qin, J. Shan, and M. Lu, "Panoramic slam from a multiple fisheye camera rig," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 159, pp. 169–183, 2020.

[7] A. Chang, A. Dai, T. Funkhouser, M. Halber, M. Niessner, M. Savva, S. Song, A. Zeng, and Y. Zhang, "Matterport3d: Learning from rgb-d data in indoor environments," *arXiv preprint arXiv:1709.06158*, 2017.

[8] M. Rey-Area, M. Yuan, and C. Richardt, "360monodepth: High-resolution 360deg monocular depth estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 3762–3772.

[9] I. Armeni, S. Sax, A. R. Zamir, and S. Savarese, "Joint 2d-3d-semantic data for indoor scene understanding," *arXiv preprint arXiv:1702.01105*, 2017.

[10] S.-H. Chou, C. Sun, W.-Y. Chang, W.-T. Hsu, M. Sun, and J. Fu, "360-indoor: Towards learning real-world objects in 360deg indoor equirectangular images," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2020, pp. 845–853.

[11] G. Pandey, J. R. McBride, and R. M. Eustice, "Ford campus vision and lidar data set," *The International Journal of Robotics Research*, vol. 30, no. 13, pp. 1543–1552, 2011.

[12] Z. Javed and G.-W. Kim, "Panovild: a challenging panoramic vision, inertial and lidar dataset for simultaneous localization and mapping," *The Journal of Supercomputing*, vol. 78, no. 6, pp. 8247–8267, 2022.

[13] J. Liu, D. Liu, W. Yang, S. Xia, X. Zhang, and Y. Dai, "A comprehensive benchmark for single image compression artifact reduction," *IEEE Transactions on image processing*, vol. 29, pp. 7845–7860, 2020.

[14] W. Yang, Y. Qian, J.-K. Kämäräinen, F. Cricri, and L. Fan, "Object detection in equirectangular panorama," in *2018 24th international conference on pattern recognition (icpr)*. IEEE, 2018, pp. 2190–2195.

[15] A. Petrovai and S. Nedevschi, "Semantic cameras for 360-degree environment perception in automated urban driving," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 10, pp. 17 271–17 283, 2022.

[16] L. Yang, B. Kang, Z. Huang, X. Xu, J. Feng, and H. Zhao, "Depth anything: Unleashing the power of large-scale unlabeled data," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 10 371–10 381.

[17] D. Tu, H. Cui, and S. Shen, "Panovlm: Low-cost and accurate panoramic vision and lidar fused mapping," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 206, pp. 149–167, 2023.

[18] K. Yuan, L. Ding, M. Abdelfattah, and Z. J. Wang, "Licas3: A simple lidar–camera self-supervised synchronization method," *IEEE Transactions on Robotics*, vol. 38, no. 5, pp. 3203–3218, 2022.

[19] B. Berenguel-Baeta, J. Bermudez-Cameo, and J. J. Guerrero, "Fredsnet: Joint monocular depth and semantic segmentation with fast fourier convolutions from single panoramas," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 6080–6086.

[20] Y. Liao, J. Xie, and A. Geiger, "Kitti-360: A novel dataset and benchmarks for urban scene understanding in 2d and 3d," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 3, pp. 3292–3310, 2022.

[21] Q. Feng, H. P. Shum, and S. Morishima, "360 depth estimation in the wild-the depth360 dataset and the segfuse network," in *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2022, pp. 664–673.

[22] G. Kim, S. Yun, J. Kim, and A. Kim, "Sc-lidar-slam: A front-end agnostic versatile lidar slam system," in *2022 International Conference on Electronics, Information, and Communication (ICEIC)*, 2022, pp. 1–6.

[23] P. Koschorrek, T. Piccini, P. Öberg, M. Felsberg, L. Nielsen, and R. Mester, "A multi-sensor traffic scene dataset with omnidirectional video," in *2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2013, pp. 727–734.

[24] Y. C. Huajian Huang, Yinzhe Xu and S.-K. Yeung, "360vot: A new benchmark dataset for omnidirectional visual object tracking," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2023.

[25] H. Huang and S.-K. Yeung, "360vo: Visual odometry using a single 360 camera," in *International Conference on Robotics and Automation (ICRA)*. IEEE, 2022.

[26] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "nuscenes: A multimodal dataset for autonomous driving," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 11 621–11 631.

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

[28] S. Yogamani, C. Hughes, J. Horgan, G. Sistu, P. Varley, D. O'Dea, M. Uricár, S. Milz, M. Simon, K. Amende *et al.*, "Woodscape: A multi-task, multi-camera fisheye dataset for autonomous driving," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 9308–9318.

[29] N. Carlevaris-Bianco, A. K. Ushani, and R. M. Eustice, "University of michigan north campus long-term vision and lidar dataset," *The International Journal of Robotics Research*, vol. 35, no. 9, pp. 1023–1035, 2016.

[30] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, A. Y. Ng *et al.*, "Ros: an open-source robot operating system," in *ICRA workshop on open source software*, vol. 3, no. 3.2. Kobe, Japan, 2009, p. 5.

[31] J. Kannala and S. S. Brandt, "A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses," *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 8, pp. 1335–1340, 2006.

[32] J. Beltrán, C. Guindel, A. de la Escalera, and F. García, "Automatic extrinsic calibration method for lidar and camera sensor setups," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 10, pp. 17 677–17 689, 2022.

[33] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and M. J. Marín-Jiménez, "Automatic generation and detection of highly reliable fiducial markers under occlusion," *Pattern Recognition*, vol. 47, no. 6, pp. 2280–2292, 2014.

[34] P. Furgale, J. Rehder, and R. Siegwart, "Unified temporal and spatial calibration for multi-sensor systems," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2013, pp. 1280–1286.

[35] C. Sun, M. Sun, and H.-T. Chen, "Hohonet: 360 indoor holistic understanding with latent horizontal features," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 2573–2582.

[36] 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 Transactions on Robotics*, vol. 37, no. 6, pp. 1874–1890, 2021.

[37] Z. Zhang and D. Scaramuzza, "A tutorial on quantitative trajectory evaluation for visual(-inertial) odometry," in *IEEE/RSJ Int. Conf. Intell. Robot. Syst. (IROS)*, 2018.

[38] I. Kasahara, S. Agrawal, S. Engin, N. Chavan-Dafle, S. Song, and V. Isler, "Rico: Rotate-inpaint-complete for generalizable scene reconstruction," *arXiv preprint arXiv:2307.11932*, 2023.

[39] A. D. Vuong, M. N. Vu, B. Huang, N. Nguyen, H. Le, T. Vo, and A. Nguyen, "Language-driven grasp detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 17 902–17 912.