2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 14-18, 2024. Abu Dhabi, UAE

SwiftBase: A Ddataset based on High-Frequency Visual Measurement for Visual-Inertial Localization in High-Speed Motion Scenes

Zhenghao Zou, Yang Lyu*, Chunhui Zhao, Xirui Kao, Jiangbo Liu and Haochen Chai

***Abstract*— Localizing an aggressively moving platform is a considerable challenge in the SLAM domain. This paper presents a dataset, *SwiftBase*, crafted to facilitate research into precise localization under such conditions. It includes high-speed cameras with over 200Hz sampling rate, capturing detailed visual data for analyzing rapid external dynamics. The dataset features two IDS high-speed cameras, a low-frequency camera, and a high-precision integrated inertial measurement unit (IMU). Calibration parameters are provided, and sensor data is synchronized using ROS system time. *SwiftBase* is recorded in indoor environments, utilizing pulleys and suspension ropes to simulate high-speed conditions, with ground truth data supplied by OptiTrack. *SwiftBase* has been instrumental in evaluating advanced VI-SLAM algorithms. However, there is still an urgent need for new algorithms capable of robust and real-time tracking in High-Speed localization.**[1]

I. INTRODUCTION

The burgeoning field of competitive drone racing seeks to push the boundaries and foster the development of autonomous sensing, planning, and state estimation technologies for drones in high-speed and complex environments[1][2]. Capable of achieving remarkable speed and agility, racing drones can swiftly navigate intricate environments and cover extensive ground areas with minimal time expenditure. Such capabilities hold significant potential for a wide range of applications, including emergency rescue, exploration of unknown territories, autonomous driving, and extreme testing of platforms[3][4].

The ability of racing drones to move autonomously and navigate with precision hinges on environmental perception and state estimation. Traditional Simultaneous Localization and Mapping (SLAM) technology actively performs tracking and mapping. By providing fundamental environmental and positional data for the drone, SLAM technology lays the groundwork for effective planning and decision-making.

With the maturation and broadening applications of 3D mobile robot technology, sensors based on vision and lidar have gained widespread use, leading to the development of Visual and Laser SLAM algorithms[5][6], [7]. While laser-based SLAM provides a wide field of view, simultaneous ranging and speed measurements, it suffers from drawbacks such as sparse scene representation, high cost, and unsuitability for airborne use. Conversely, visual SLAM which utilizes lightweight optical components, captures complex scene textures and aligns with human perceptual logic, offering the advantage of semantic information integration. However, single sensors are not always effective in complex environments. Visual-Inertial SLAM (VI-SLAM), which combines visual measurements to correct IMU drift errors and uses IMU data for a priori motion information, is considered the most compact yet complete system[8], providing enhanced accuracy and robustness[9], [10].

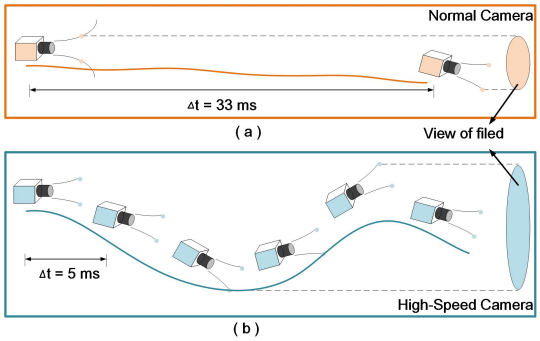


Fig. 1. Unique advantages of high-speed cameras. (a) A low frame rate camera operates at 30Hz. (b) A high frame rate camera operates at 200Hz. The increased sampling frequency of the high-speed camera allows for the observation of more detailed states and a broader field of view, resulting in an estimated trajectory that is closer to the actual movement trajectory.

In the challenges of rapid positioning, VI-SLAM represents a reliable approach within the spectrum of current methodologies. The prevalent use of low frame rate imaging leads to blurred images, excessive disparity, and a discrepancy between perception and movement due to swift motion, all of which restrict a robot's capacity for high-mobility tasks. As proposed by [24], employing high-speed cameras can provide high frame rate images, making the state observable in process. Illustrated in Fig 1, the high-speed camera operates at 200Hz, matching the IMU's observation frequency and ensuring consistency between perception and motion. Additionally, the abundance of images refines the estimated trajectory, aligning it closer with the actual movement and enhancing accuracy in uncharted environments positing.

This article introduces a research dataset for high-speed motion SLAM, focused on indoor environments and presents a high-speed perception system. This dataset captures sensor data during intricate high-speed motion processes, sufficient for reconstructing high-speed motion trajectories. Although the best-performing VI-SLAM algorithm on this dataset was evaluated, no existing VI-SLAM algorithm can validly

Zhenghao Zou, Chunhui Zhao, Xirui Kao, Jiangbo Liu, Haochen Chai Yang Lyu are with the School of Automation, Northwestern Polytechnical University, Xi'an Shanxi, 710072 China. email: lyu.yang@nwpu.edu.cn.
This work was supported by the National Natural Science Foundation of China under Grant 62203358, 62233014, and 62073264.

[1]This project is open at: https://github.com/ZzhYgwh/High-Speed-Motion-Dataset

TABLE I

OVERVIEW OF VISUAL-INERTIAL SLAM DATASET

Dataset	Scenes	Type	Camera	IMU	Lidar	UWB	Event	Infrared	GT	Camera Fre.(Hz)
Common motion scene										
KITTI[11]	Urban	Car	✓	✓	✓				✓	10
EuRoC Mav[12]	Indoor	UAV	✓	✓					✓	20
TUM VI[13]	Indoor/Outdoors	Handle-held	✓	✓					✓	30
Zurich Urban MAV[14]	Urban	UAV	✓	✓					✓	30
KAIST[15]	Urban	Car	✓	✓	✓				✓	-
DeepIO[16]	Indoor	Handle-held	✓	✓					✓	100
M2DGR[17]	Indoor/Outdoors	Car	✓	✓	✓		✓	✓	✓	15
NTU VIRAL[18]	Indoor/Outdoors	UAV	✓	✓	✓	✓			✓	10
High-speed motion scene										
UZH-FPV[19]	Indoor/Outdoors	UAV	✓	✓			✓		✓	30
RACECAR[20]	Race	Car	✓	✓	✓				✓	8
Roller-Coaster[21]	Roller	Handle-held	✓	✓						30
TII-RATM[22]	Indoor	UAV	✓	✓					✓	120
Blackbird[23]	Render Env	UAV	✓	✓					✓	120
SwiftBase(ours)	**Indoor**	**UAV**	✓	✓					✓	**30 / 200**

achieve robust and real-time tracking. Our contributions are as follows:

• We have constructed high-speed sensing systems that offer devices and drivers capable of 200Hz high-frequency visual-inertial measurements. The system is outfitted with dual high-speed cameras, a low frame rate camera, and an IMU. We provide calibration results and platform device drivers.

• We have designed experimental scenarios and constructed high-speed motion SLAM datasets. The scenario is designed in indoor environments, with ground truth provided by Optitrack. All data sequence and calibration results are made available. We simulated complex high-speed motion patterns and recorded sensor data in this scenario. To the best of our knowledge, this is the first dataset based on high-speed camera for SLAM applications.

• We have established a benchmark for high-speed localization issues. Various VIO or VI-SLAM were employed as participants, with the Absolute Pose Error (APE), Relative Pose Error (RPE) evaluated by evo and the overall algorithm's runtime index serving as experimental benchmarks. Ultimately, the dataset's best-performing VI-SLAM algorithm was assessed, and the characteristics of high-speed positioning were delineated.

II. RELATED WORKS

Several classic VI-SLAM datasets, such as KITTI[11], EuRoC Mav[12], and TUM VI[13], have emerged as benchmark data for numerous VI-SLAM algorithms. The Kitti dataset, which relies on an outdoor vehicle-mounted platform equipped with a variety of sensors, provides benchmarks for urban road scenarios, including odometry, road detection, scene flow, and more. Meanwhile, it also presents challenges such as lighting variations, dynamic object interference, and occlusion. EuRoC Mav dataset poses indoor challenges to VI-SLAM, with its unbalanced lighting and complex motion. Simultaneously, its high frame rate ground truth data allows for a detailed comparison of VI-SLAM performance. TUM-VI explores the impact of changing illumination and low-texture areas in various types of environments, including indoors, corridors, halls, pipes, and outdoors.

In recent years, the Zurich Urban MAV[14] has utilized micro-aerial vehicles to conduct low-altitude street-view aerial photography of towns. Its changing perspective, lighting variations, and repeated textures also necessitate special consideration in VI-SLAM. The KAIST[15] dataset, recorded both during the day and night, provides a foundation for visual loop closure detection in SLAM. The DeepIO[16] dataset supports the development of deep learning in VI-SLAM, including a wide range of generalizable scenes and motion modes, and supports data generation and multi-type algorithm evaluation channels. The M2DGR[17] dataset provides a large multi-sensor integrated ground robot SLAM dataset, including fisheye camera, infrared camera, event camera, lidar, RTK, and an independent VI device. The NTU VIRAL[18] dataset offers a new benchmark for autonomous driving, providing multiple types of sensors such as cameras, radars, IMUs, and UWB, with the aim of providing a high-precision positioning dataset.

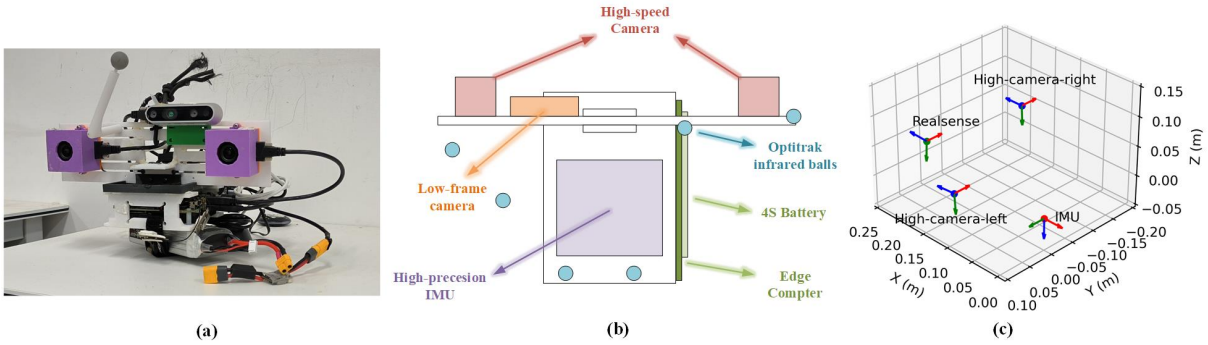
These datasets strive to enhance the sensor accuracy and scene generalization capabilities of SLAM datasets, thereby propelling the advancement of VI-SLAM technology. However, some current extreme scenarios cannot be validated using the aforementioned datasets. High-Speed scenarios is one of the challenges that SLAM currently needs to address. Particularly in the field of autonomous driving, achieving fast and robust positioning under high-speed conditions has become a significant challenge.

Currently, several studies have begun to concentrate on the SLAM problem in high-speed motion scenarios. We summarize the VI-SLAM datasets of traditional motion and high-speed motion scenes in Table I. The UZH-FPV dataset[19] offers high-speed flight data. The UZH-FPV dataset can validate the speed and robustness of VI-SLAM under high-speed motion. RACECAR[20] introduces a fully automated racing high-speed dataset. This dataset provides environmental data for autonomous driving under extreme speed conditions. A set of high-speed video sequences recorded on the Shanghai Disneyland roller coaster has been released on GitHub[21]. The data sequence includes a set of binocular stereo vision and IMU equipment, with a top speed of 98km/h. This data poses challenges to the VIO algorithm under high-speed

TABLE II

PLATFORM AND SYSTEM FOR DATA ACQUISITION

Device	System	Driver	Num	Frequency(Hz)	Description
		Sensor			
IDS U3-3041LE Rev.1.2	ROS	ids_ros_driver	2	200	High-speed Camera
Parker 3DMGQ7-GNSS/INS	ROS	microstrain_inertial_driver	1	200	High-precision IMU
Inter Realsense D435i	ROS	realsense-ros	1	30	Realsense D435i
		Platform			
Intel NUC	Ubuntu 20.04	ROS noetic	1	-	Drive devices, Collect and Store dataset
	Optitrack		1	360	Provide groud truth
	Motion simulation platform		1	-	Simulate high-speed translation and rotation



(a) (b) (c)

Fig. 2. High-Speed Sensing Platform: (a) illustrates the installed equipment platform. (b) depicts the structural information of the platform and the layout of the sensors. (c) demonstrates the relationship between the external parameters of the sensors after calibration using Kalibr [22][23][25][26][27]. The high-speed perception platform comprises two high-speed cameras, a low frame rate camera, a high-precision IMU, an Inter NUC edge computer, some essential structural fixtures and power supply equipment. All cameras are primarily located on the front sliding plate, while the IMU is situated behind the platform. The NUC computer is installed in a semi-enclosed space beneath the platform, and the 4S battery is positioned at the bottom.

motion and low illumination conditions. TII-RATM[22] and Blackbird[23] both provide 120Hz visual images in actual UAV aggressive flight and visual rendering environment.

These studies explore the challenges of visual inertial SLAM in high-speed motion, particularly the state estimation problem under high-speed motion. However, these studies still rely on low-frequency visual measurements, relative to the common low-cost IMU measurement frequency (200Hz), for attitude estimation under high-speed motion. Although IMU data can alleviate this problem in the short term, it will reduce the accuracy and robustness of SLAM algorithm. Therefore, the *SwiftBase* proposed in this article adopts a new high-speed camera and combines motion data provided by IMU for visual measurement. Our dataset aims to provide a suitable perception platform for high-speed motion scenes, while revealing the potential localization conflicts that may arise when inertial data and visual data are on the same frequency.

III. SWIFTBASE

A. Sensor Setup

We have developed a high-speed perception system tailored for state estimation in high-maneuver scenarios, as depicted in Fig 2.(a). The vision sensor employed is the IDS U3-3041LE Rev.1.2, boasting a maximum frame rate of 230 FPS and a resolution of 1456 x 1088 pixels. It is crucial to acknowledge that the elevated frame rate of this industrial camera reduces exposure time, leading to diminished image intensity indoors. To address this, we recommend the implementation of image enhancement algorithms to augment the quality of high-speed images, such as histogram equalization algorithm. The operating frequency of high-speed cameras remains at 200Hz, which is comparable to the widely used low-cost IMU measurement frequency. Besides, this paper specially proposes the challenging problem when motion inertial information and visual information are on the same frequencyFurthermore, our system includes an Intel Realsense D435i camera, operating at a lower frame rate of 30Hz, which serves as a benchmark for the dataset's aggressive motion, thereby demonstrating the distinct advantage of high-speed cameras. The chosen IMU is the Parker 3DMGQ7-GNSS/INS, a high-precision device capable of integrating GNSS data through a filter fusion algorithm, enabling centimeter-level localization. Nonetheless, our dataset solely utilizes the IMU data, which is set to 200 Hz to assess VI-SLAM performance.

All sensors are mounted on custom-designed 3D structures, as illustrated in Fig 2.(b). The high-speed cameras are aligned parallel to the platform's front-end, with a low-frame-rate camera positioned adjacent to the left high-speed camera. The IMU is strategically placed in the center rear to maintain the platform's balance. Comprehensive details of all equipment, drivers, and operational characteristics of the high-speed sensing system are encapsulated in Table II.

B. Platform

In addition to the high-speed perception system, the platform-level auxiliary equipment includes an Intel NUC edge computer with high broadband transmission capabilities, a personal computer, pulleys and slings, a 4S battery, and necessary fixing materials. The NUC is equipped with the Ubuntu 20.04 operating system, ROS Noetic and essential drivers, which are utilized to operate sensors, synchronize

TABLE III

SWIFTBASE

Sequence	Duration(s)	Dist(m)	ROSbag	Translate	Rotation
low_trans	90	23.218	low_trans.bag	Low	none
low_rot	113	12.485	low_rot.bag	none	Low
low_comb	122	22.333	low_comb.bag	Low	Low
mid_trans	56.6	16.331	mid_trans.bag	Middle	none
mid_rot	106	12.552	mid_rot.bag	none	Middle
mid_trans_low_rot	59	7.603	mid_trans_low_rot.bag	Middle	Low
low_trans_mid_rot	82	13.949	low_trans_mid_rot.bag	Low	Middle
mid_comb	66	12.219	mid_comb.bag	Middle	Middle
high_trans	79	11.185	high_trans.bag	High	none
high_rot	111	16.618	high_rot.bag	none	High
high_trans_low_rot	62	10.690	high_trans_low_rot.bag	High	Low
low_trans_high_rot	77	16.010	low_trans_high_rot.bag	Low	High
high_trans_mid_rot	56.9	16.020	high_trans_mid_rot.bag	High	Middle
mid_trans_high_rot	89	17.543	mid_trans_high_rot.bag	Middle	High
high_comb	96	20.386	high_comb.bag	High	High
loop_mid	135	84.316	loop_mid.bag	Middle	Middle

acquisition times with the ROS system time, provide external auxiliary power, and store data. The PC is configured with ROS Noetic, Kalibr[25][26][27][28][29], imu_util[30], and the environmental dependencies for the benchmark algorithm in section IV-B. These configurations are employed to calibrate the platform, organize datasets, and conduct algorithm evaluation experiments separately. The native configuration of the PC includes a 6-core Intel i5-11400H processor with 16GB of memory.

C. Calibration

The dataset includes a 50-second ROS bag for calibration, along with the corresponding calibration results. We have adjusted the high-speed camera's sampling frequency to 4 Hz and set the exposure time to 30 ms to ensure clearer time alignment and sharper images.

Our calibration process is structured into four stages. In the first stage, Kalibr is employed to perform pinhole model parameters calibration for each camera, including both high-speed and low frame rate cameras. In the second stage, imu_util is used to analyze the Allan variance of the IMU data to determine the bias and random walk of the IMU's gyroscope and accelerometer. In the third stage, Kalibr calibrates the external parameters of each camera and IMU. Finally, we use IMU as a reference to delineate the complete external parameter relationship, as illustrated in Fig 2.(c).

We refrained from directly matching the parameters obtained between high-speed cameras or between a high-speed camera and a low-frame-rate camera. This is because camera images may vary, and separate calibration allows us to capture the unique characteristics of each camera, thereby minimizing the introduction of model errors as much as possible.

D. Motion

In VSLAM, rotational motion causes more significant visual changes than transition, making high-speed rotation a challenge. In rotor drones, attitude and control are coupled, with the X, Y, and YAW channels being more sensitive. To simulate actual flight scenarios, we design experimental conditions that include rotation around the Z-axis on a horizontal plane.

During the dataset collection process, the movement conditions of the platform were strictly controlled. The platform

(a) high-speed seq (b) high-speed seq with Histogram equalization (c) Realsense D435i

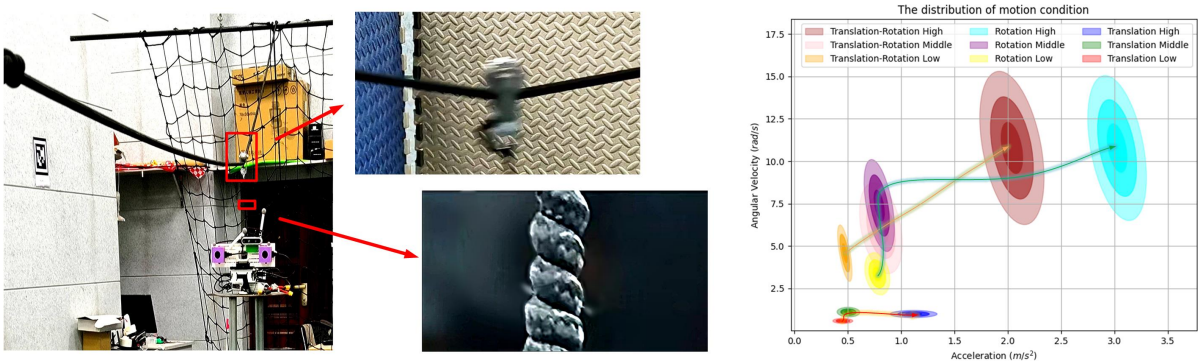
Fig. 3. Partial sequence images. Partial sequences taken from low_comb, mid_comb and high_comb. The original image of the high-speed sequence (a) has weak intensity, and histogram equalization is used to improve the image contrast (b). Finally, a low frame rate image (c) is shown. It can be shown that after the images collected by high-speed cameras are adjusted, the image quality and clarity are less affected by violent motion.

was set to slide from three different heights on the zipline, creating varying degrees of acceleration. Different angular velocities were imparted to the traction platform due to the varying extent of distortion in the suspension rope. For a more nuanced comparison, high-speed motion was categorized into distinct translational and rotational motions, as well as combined motions, to facilitate the analysis of motion characteristics under high-speed conditions.

To illustrate the incremental boundary conditions for the designed experimental motion, the maximum acceleration and angular velocity observed in each sequence were used as the criteria for differentiation. Acceleration and angular velocity information was collected from the IMU data of each sequence, with the acceleration being corrected for gravity compensation. Approximately, the experimental motion decision boundary was defined as follows: For the platform's three-axis composite angular velocity, Ω, velocities of $\Omega \leq 6$ rad/s are considered low-speed rotation, velocities within 6 rad/s $< \Omega \leq 9$ rad/s are medium-speed rotation, and velocities of $\Omega > 9$ rad/s are high-speed rotation. After compensating for gravitational acceleration, the three-axis composite acceleration, a, is minimal; thus, the maximum acceleration is used to distinguish translational motion. An acceleration of $a \leq 0.4$ m/s^2 indicates low-speed translation, 0.4 m/s^2 $< a \leq 1.0$ m/s^2 signifies medium-speed translation, and $a > 1.0$ m/s^2 defines high-speed translation.

E. Data Collection

Upon the commencement of the data collection, it is observed that the complexity of the proposed dataset lies not in the scene but in the motion. The data highlighted the complexity and challenges of high-speed motion, so only indoor experiments were designed. The indoor scene is a research laboratory equipped with Optitrack, a high-precision visual positioning system. The complexity of actions is based on the standards of section III-D, and the



(a) lab environment (b) motion driver (c) motion distribution

Fig. 4. High speed motion experiment. (a) shows the experimental scene of the high-speed motion experiment, using pulleys and suspension ropes to connect the sensing platform. (b) shows that in the actual experiment, the pulley motion provides linear acceleration, and the deformation of the suspension rope is the cause of the angular velocity. (c) shows the distribution and statistics of acceleration and angular velocity under different motions analysed by IMU data (3σ area).

datasets recorded are arranged according to the methods of individual complex actions and combined complex actions. For instance, high-speed rotation and high-speed translation are decomposed into three data packets, including individual high-speed rotation, high-speed translation, and combined high-speed action. This approach provides researchers with more detailed references and comparisons, emphasizing the characteristics of high-speed, complex motion.

It is important to note that the high frame rate of high-speed images results in a surge in storage capacity. The *sensor_msgs/CompressedImage* message type in ROS is utilized to compress and encode images, making the data concise and easy to download and use. Upon completion of data collection, necessary deletions are performed on the data, including trimming the informal part of the sequence, eliminating time-reverse data, and image decompression and adjustment. Each data sequence provides an initialization process for algorithm execution to ensure a robust initial state.

Fig 4 illustrates the actual experimental scene constructed, the motion driving method, and the analysis of motion data. Pulleys are used to generate linear acceleration indoors, and the torsional deformation of hanging ropes is used to generate angular velocity. The sensing platform is balanced by the center of gravity and driven by pulleys and hanging ropes to collect data. All data sequences are listed in Table III, and the characteristics and composition of the data sequences are described in detail. A set of examples is also provided to illustrate the acquisition effect of high-speed motion, as shown in Fig 3. Based on the above analysis and interpretation, the dataset is comprehensive for high-speed motion measurements and provides all available replication materials.

IV. Experiment Evaluation

A. Aggression assessment

In this section, the focus is on elucidating the degree of motion of the *SwiftBase* collection platform and the unique contribution of high-speed cameras to high-speed motion. In the evaluation of motion within the dataset, IMU measurements have been utilized by researchers previously, as it can provide the acceleration and angular velocity of the body. However, this representation is subject to controversy. Firstly, the acceleration and angular velocity of the body are high-order movements and do not provide an intuitive movement reference. Even when the IMU pre-integration method is employed, the estimation results are prone to being affected by noise bias and gravity interference, rendering them unreliable. Secondly, existing motion methods often lack independent rotation and translation, but often appear in combination, leading to ambiguity about the degree of movement.

Aggressive motion prompts the camera to traverse a wide range of angles, resulting in significant differences in the sequence of consecutive images. The optical flow of the image can accentuate this difference. The focal length normalized optical flow speed proposed in [19] is adopted to measure the intensity of actual motion. In accordance with [19], dense optical flow is estimated using an OpenCV implementation of Farneback's method. After eliminating some optical flows with small values, the optical flow is normalized according to the image resolution, and then normalized according to the camera focal length. This evaluation method is applied to the faster datasets or sequences EuRoC (V1_03, V2_03), UZH FPV (indoor, outdoor) and the data sequences low_comb, mid_comb and high_comb of this article as dataset evaluation criteria. Subsequently, the low frame rate camera images and high-speed camera images of this paper's dataset were evaluated to compare the advantages of high-speed cameras and traditional cameras. Finally, the sampling frequency of the high-speed camera is adjusted, aiming to compare the impact of high-speed camera frequency on aggressive motion estimation.

The experimental results are depicted in Fig 5. Among them, Fig 5.(a) compares the focal length-normalized optical flow velocity of different datasets, which can illustrate that the acquisition conditions of *SwiftBase* have higher dynamic characteristics. Fig 5.(a) also illustrates that in the movement standards collected by the dataset designed above, different degrees of high-speed movement are indeed

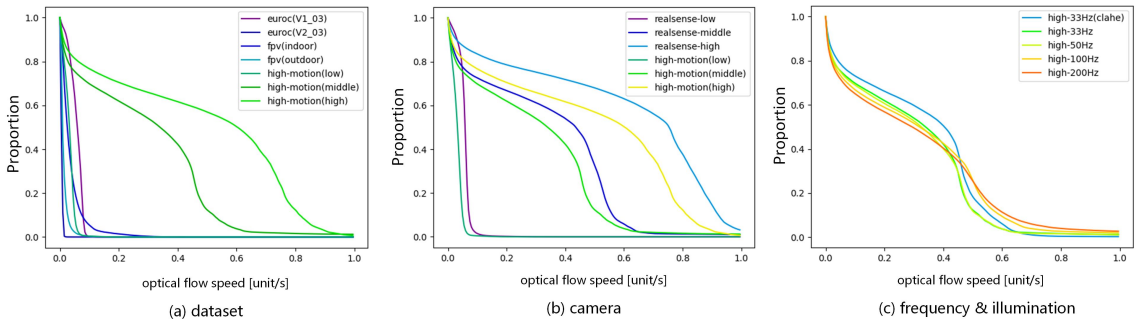


Fig. 5. Comparison of dynamic range of normalized focal length optical flow velocity. (a) shows the normalized optical flow proportions of some data sequences of EuRoC, UZH FPV and *SwiftBase*. *SwiftBase* captures high-speed motion and has higher dynamics. (b) shows the focal length-normalized optical flow velocity ratio between low frame rate cameras and high-speed cameras. It shows that high-speed cameras can reduce image differences and improve the quality of images in high-speed motion. (c) shows the normalized optical flow velocity ratio of high-speed images at different sampling frequencies. The higher the sampling frequency, the smaller the change rate of the optical flow proportion as the optical flow speed increases. The optical flow of high-frequency images tends to be uniform, which can ensure the image tracking effect.

distinguished. Fig 5.(b) shows that in the same data sequence, the normalized optical flow distribution of the high-speed camera is more uniform, which can provide clearer visual images and reduce the motion difference between images. Fig 5.(c) shows that higher sampling frequency in high-speed cameras also provides smoother optical flow distribution, which shows that high frame rate sampling can improve the problems of motion blur and inconsistency. Finally, a discussion on lighting conditions is proposed. In section III-A, it is shown that high-speed images are low-illumination, which has an impact on subsequent state estimation. In Fig 5.(c), it can be proven that using image enhancement algorithms to improve lighting conditions can provide optical flow capacity, which improves the robustness of state estimation. However, according to the above conclusion, this also reduces the quality of the image and weakens the accuracy of state estimation.

B. Benchmark

A comparison was conducted among various Visual-Inertial SLAM(VI-SLAM) algorithms on the *SwiftBase*, and the latest technology was evaluated on this dataset. The algorithms selected for comparison include: VINS-Mono[31], ORB-SLAM3[32], ROVIO[33], OpenVINS[34], DM-VIO[35] and EqVIO[36]. To ensure the compatibility of the algorithm with this dataset, necessary modifications were made, which included the time measurement process, image decompression operation and updating the calibration parameter file. Apart from these, no other design processes and strategies of the algorithm were modified. At the commencement of the experiments, the data intervals were intentionally spanned to ensure that all SLAM algorithms had robust initial state estimates. Loop detection and relocation capabilities were retained for all tested algorithms, guaranteeing the original design of the algorithm and unbiased estimation results.

In the experiments, the focus was on three metrics: absolute pose error (APE), relative pose error (RPE) and estimated overall time. APE reflects the global estimation accuracy of the algorithm, RPE shows the local estimation accuracy of the algorithm, and the overall estimation time reflects the estimation speed of the algorithm. Based on these two indicators, the performance of the high-speed motion algorithm was verified, and the visual-inertia estimation algorithm that takes into account both accuracy and speed was evaluated.

C. SOTA Algorithm

Certain traditional frequency-based VI-SLAM algorithms such as ORB-SLAM3 and DM-VIO are unable to track high-speed data effectively. OpenVINS, either fails to initialize successfully or loses positioning during tracking. As per the evaluation results in Table IV and the positioning trajectory results in Fig 6, the overall test algorithm exhibits varying degrees of positioning drift, and the positioning accuracy does not reach the centimeter level. From the perspective of a single algorithm, ROVIO outperforms other algorithms in terms of positioning accuracy and estimation time, making it the best algorithm tested on the dataset to date. We provide the performance of the best algorithm ROVIO on different datasets, as shown in Fig 7, to compare the potential advantages of using high-speed measurement.

The reasons why traditional VI-SLAM algorithms cannot position accurately and quickly are summarized as follows: (i) Insufficient IMU pre-integration. The frequency ratio between high-speed images and IMU data is close to 1, indicating that the IMU pre-integration process is too short to generate sufficient motion prior information. (ii) The movement speed is too fast and the amplitude of the IMU single frame data is too large, which affects the consistency of vision and IMU constraints. (iii) The processing speed of the algorithm still cannot reach real-time performance, and there is serious data lag, resulting in further positioning drift. However, our dataset guarantees the tracking quality of images under high-speed motion, which has significant advantages for high-speed positioning.

Therefore, to achieve rapid positioning, some dedicated SLAM algorithms based on high-speed data need to be redesigned and developed. We plan to make further improvements in the future. We also look forward to researchers interested in this work communicating with us and proposing algorithms with better performance.

TABLE IV

PERFORMANCE ON SWIFTBASE

Methods	low_trans			low_rot			low_comb			mid_trans			mid_rot		
	APE(m)	RPE(m)	Time(s)	APE(m)	RPE(m)	Time(s)	APE(m)	RPE(m)	Time(s)	APE(m)	RPE(m)	Time(s)	APE(m)	RPE(m)	Time(s)
VINS-Mono	**1.104**	0.227	114.586	-	-	-	-	-	-	1.005	-	136.534	-	-	-
ORB-SLAM3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
DM-VIO	-	-	119.690	-	-	-	-	-	-	-	-	-	-	-	-
OpenVINS	-	-	-	-	-	-	-	-	-	0.721	-	-	-	-	-
ROVIO	1.108	0.004	**108.293**	**0.069**	**0.004**	**164.985**	0.953	**0.006**	**140.038**	**0.257**	**0.006**	**69.796**	-	-	-
EqVIO	1.230	**0.003**	542.371	0.075	0.006	714.508	**0.405**	0.009	773.267	1.032	0.008	325.003	-	-	-
	mid_trans_low_rot			low_trans_mid_rot			mid_comb			high_trans			high_rot		
	APE(m)	RPE(m)	Time(s)	APE(m)	RPE(m)	Time(s)	APE(m)	RPE(m)	Time(s)	APE(m)	RPE(m)	Time(s)	APE(m)	RPE(m)	Time(s)
VINS-Mono	1.177	0.036	233.148	-	-	-	-	-	-	0.515	0.182	166.359	-	-	-
ORB-SLAM3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
DM-VIO	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
OpenVINS	0.824	0.251	-	-	-	-	1.252	0.043	-	1.937	0.095	-	-	-	-
ROVIO	**0.248**	**0.006**	**69.518**	**0.672**	**0.006**	**93.844**	1.847	**0.006**	**71.200**	**0.513513**	**0.005**	**95.841**	-	-	-
EqVIO	1.161	1.161	650.503	0.880	0.704	479.705	**1.145**	0.007	392.722	1.194	0.015	450.931	-	-	-
	high_trans_low_rot			low_trans_high_rot			high_trans_mid_rot			mid_trans_high_rot			high_comb		
	APE(m)	RPE(m)	Time(s)	APE(m)	RPE(m)	Time(s)	APE(m)	RPE(m)	Time(s)	APE(m)	RPE(m)	Time(s)	APE(m)	RPE(m)	Time(s)
VINS-Mono	-		-	-	-	-	0.437	-	103.689	-	-	-	-	-	-
ORB-SLAM3	-		-	-	-	-	-	-	-	-	-	-	-	-	-
DM-VIO	-		-	-	-	-	-	-	-	-	-	-	-	-	-
OpenVINS	-		-	-	-	-	0.610	-	-	-	-	-	-	-	-
ROVIO	**0.563**	0.007	**74.901**	**0.752**	0.008	**94.521**	**0.426**	1.671	**63.065**	-	-	-	-	-	-
EqVIO	1.241	**0.006**	357.656	0.938	**0.007**	449.400	0.506	**0.012**	299.845	**0.797**	**0.018**	**523.736**	**0.686**	**0.686**	**587.177**

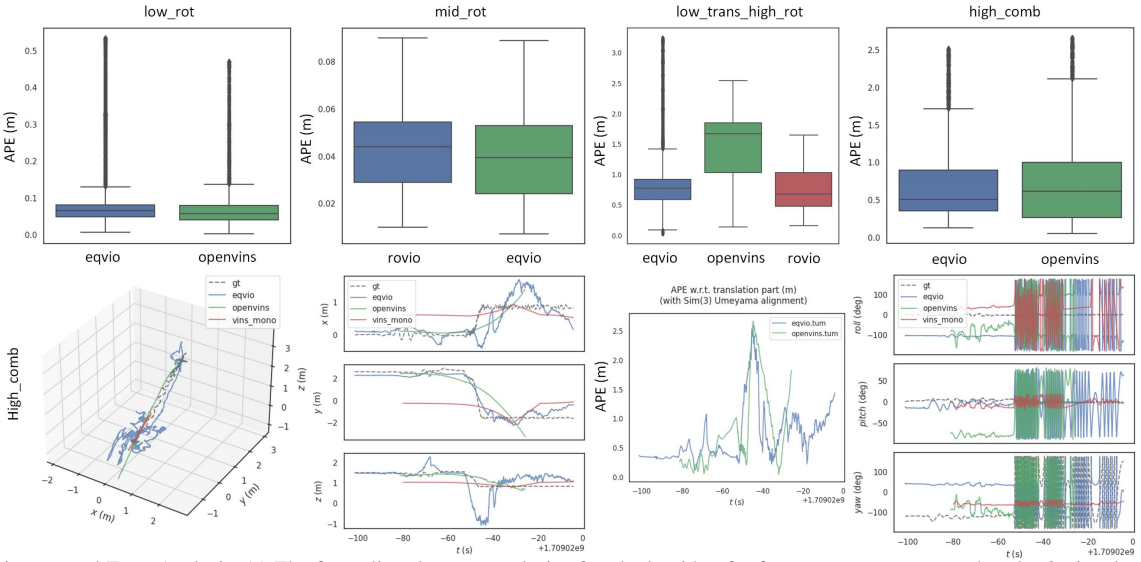


Fig. 6. Trajectory and Error Analysis. (a) The figure lists the error analysis of each algorithm for four sequences, compared to the Optitrack ground truth. (b) The error analysis details are listed separately for high-speed rotation and high-speed translation. High-speed rotation poses a significant challenge to visual localization. High frame rate images make it difficult for traditional visual localization algorithms to track correctly. Insufficient IMU pre-integration leads to further localization drift.

D. Pressure Test

High-speed measurement enhances the perception of the external world, but the substantial number of visual images generated poses a significant load for SLAM algorithms. The data processing speed of several algorithms, including VINS-Mono, ROVIO, and EqVIO, was evaluated. However, further testing details for other algorithms could not be provided as this would involve modifying the internal framework of the algorithm. Testing these algorithms requires time to test conditions that satisfy all visual data usage conditions. Take the entire process time of data-driven startup as the overall running time of the algorithm. The results are collated in Table III, serving as an indication of the algorithm's data processing performance.

The results in Table III indicate that the current SLAM algorithm cannot handle high-speed data of extreme flying in real time, and high-speed data introduces delay and load pressure. To address the positioning problem of high-speed extreme flying, the development of faster algorithms is necessary. From the comparison of various algorithms, ROVIO exhibits the highest execution efficiency. This is attributed to ROVIO's use of an EKF filter architecture, which tends to have marginal computational cost.

V. CONCLUSIONS

This paper encapsulates our previously developed high-speed motion perception system and the collection of datasets. A high-speed perception system is provided, inclusive of devices and drivers. High-speed motion experiments were designed, encompassing indoor scenarios and a multitude of high-speed motion methods. It is believed that the established sensing system and dataset are comprehensive and apt for high-speed measurements. An analysis experiment on the characteristics of high-speed measurement proves that high-speed measurement has significant advantages in reducing parallax and enhancing image quality in high-speed motion. A positioning evaluation experiment was conducted on this dataset, demonstrating that ROVIO can achieve superior robust tracking on this dataset. However, it has yet

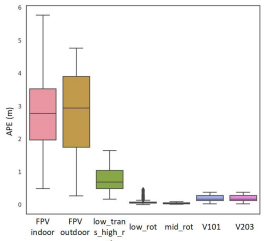


Fig. 7. The APE (Absolute Pose Error) of the ROVIO algorithm varies across datasets. In the UZH-FPV dataset, high-speed flight and low-frame-rate measurements lead to significant drift. The error is lower in the EuRoC dataset. Although high-speed visual odometry reduces errors in slow movements, it still struggles with high-speed motion, highlighting ROVIO's limitations in handling fast scenarios.

to attain the level of real-time tracking, and there remains a discrepancy in positioning accuracy. Consequently, future work will be centered on designing a SLAM framework that is suitable for high-speed tracking.

REFERENCES

[1] E. Kaufmann, L. Bauersfeld, A. Loquercio, M. Müller, V. Koltun, and D. Scaramuzza, "Champion-level drone racing using deep reinforcement learning," *Nature*, vol. 620, no. 7976, pp. 982–987, 2023.

[2] D. Hanover, A. Loquercio, L. Bauersfeld, A. Romero, R. Penicka, Y. Song, G. Cioffi, E. Kaufmann, and D. Scaramuzza, "Autonomous drone racing: A survey," *arXiv preprint arXiv:2301.01755*, 2023.

[3] Y. Song, K. Shi, R. Penicka, and D. Scaramuzza, "Reinforcement learning for agile flight: From perception to action," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023.

[4] Y. Song, K. Shi, R. Penicka, and D. Scaramuzza, "Learning perception-aware agile flight in cluttered environments," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1989–1995, IEEE, 2023.

[5] B. Zhou, Y. He, K. Qian, X. Ma, and X. Li, "S4-slam: A real-time 3d lidar slam system for ground/watersurface multi-scene outdoor applications," *Autonomous Robots*, vol. 45, pp. 77–98, 2021.

[6] M. Servières, V. Renaudin, A. Dupuis, and N. Antigny, "Visual and visual-inertial slam: State of the art, classification, and experimental benchmarking," *Journal of Sensors*, vol. 2021, pp. 1–26, 2021.

[7] S. Yi, Y. Lyu, L. Hua, Q. Pan, and C. Zhao, "Light-loam: A lightweight lidar odometry and mapping based on graph-matching," *IEEE Robotics and Automation Letters*, vol. 9, no. 4, pp. 3219–3226, 2024.

[8] A. Macario Barros, M. Michel, Y. Moline, G. Corre, and F. Carrel, "A comprehensive survey of visual slam algorithms," *Robotics*, vol. 11, no. 1, p. 24, 2022.

[9] R. Syed, S. Suriadi, M. Adams, W. Bandara, S. J. Leemans, C. Ouyang, A. H. ter Hofstede, I. van de Weerd, M. T. Wynn, and H. A. Reijers, "Robotic process automation: contemporary themes and challenges," *Computers in Industry*, vol. 115, p. 103162, 2020.

[10] Y. Lyu, S. Yuan, and L. Xie, "Structure priors aided visual-inertial navigation in building inspection tasks with auxiliary line features," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 58, no. 4, pp. 3037–3048, 2022.

[11] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The kitti dataset," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1231–1237, 2013.

[12] M. Burri, J. Nikolic, P. Gohl, T. Schneider, J. Rehder, S. Omari, M. W. Achtelik, and R. Siegwart, "The euroc micro aerial vehicle datasets," *The International Journal of Robotics Research*, vol. 35, no. 10, pp. 1157–1163, 2016.

[13] D. Schubert, T. Goll, N. Demmel, V. Usenko, J. Stückler, and D. Cremers, "The tum vi benchmark for evaluating visual-inertial odometry," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1680–1687, IEEE, 2018.

[14] A. L. Majdik, C. Till, and D. Scaramuzza, "The zurich urban micro aerial vehicle dataset," *The International Journal of Robotics Research*, vol. 36, no. 3, pp. 269–273, 2017.

[15] Y. Choi, N. Kim, S. Hwang, K. Park, J. S. Yoon, K. An, and I. S. Kweon, "Kaist multi-spectral day/night data set for autonomous and assisted driving," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 3, pp. 934–948, 2018.

[16] C. Chen, P. Zhao, C. X. Lu, W. Wang, A. Markham, and N. Trigoni, "Deep-learning-based pedestrian inertial navigation: Methods, data set, and on-device inference," *IEEE Internet of Things Journal*, vol. 7, no. 5, pp. 4431–4441, 2020.

[17] J. Yin, A. Li, T. Li, W. Yu, and D. Zou, "M2dgr: A multi-sensor and multi-scenario slam dataset for ground robots," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 2266–2273, 2021.

[18] T.-M. Nguyen, S. Yuan, M. Cao, Y. Lyu, T. H. Nguyen, and L. Xie, "Ntu viral: A visual-inertial-ranging-lidar dataset, from an aerial vehicle viewpoint," *The International Journal of Robotics Research*, vol. 41, no. 3, pp. 270–280, 2022.

[19] J. Delmerico, T. Cieslewski, H. Rebecq, M. Faessler, and D. Scaramuzza, "Are we ready for autonomous drone racing? the uzh-fpv drone racing dataset," in *2019 International Conference on Robotics and Automation (ICRA)*, pp. 6713–6719, IEEE, 2019.

[20] A. Kulkarni, J. Chrosniak, E. Ducote, F. Sauerbeck, A. Saba, U. Chirimar, J. Link, M. Behl, and M. Cellina, "Racecar-the dataset for high-speed autonomous racing," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 11458–11463, IEEE, 2023.

[21] Factor-Robotics, "Roller-coaster-slam-dataset: The world's first roller coaster slam dataset," 2023.

[22] M. Bosello, D. Aguiari, Y. Keuter, E. Pallotta, S. Kiade, G. Caminati, F. Pinzarrone, J. Halepota, J. Panerati, and G. Pau, "Race against the machine: a fully-annotated, open-design dataset of autonomous and piloted high-speed flight," *IEEE Robotics and Automation Letters*, 2024.

[23] S. Karaman, "The blackbird dataset: A large-scale dataset for uav perception in aggressive flight," in *Proceedings of the 2018 International Symposium on Experimental Robotics*, vol. 11, p. 130, Springer Nature, 2020.

[24] P. Gemeiner, A. J. Davison, and M. Vincze, "Improving localization robustness in monocular slam using a high-speed camera," in *Robotics: Science and Systems IV*, MIT Press, 2008.

[25] J. Rehder, J. Nikolic, T. Schneider, T. Hinzmann, and R. Siegwart, "Extending kalibr: Calibrating the extrinsics of multiple imus and of individual axes," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4304–4311, IEEE, 2016.

[26] 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*, pp. 1280–1286, IEEE, 2013.

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

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

[29] L. Oth, P. Furgale, L. Kneip, and R. Siegwart, "Rolling shutter camera calibration," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1360–1367, 2013.

[30] W. Gao, "imu_utils: A ros package tool to analyze the imu performance," 2018.

[31] 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.

[32] T. Qin, P. Li, and S. Shen, "Vins-mono: A robust and versatile monocular visual-inertial state estimator," *IEEE Transactions on Robotics*, vol. 34, no. 4, pp. 1004–1020, 2018.

[33] M. Bloesch, S. Omari, M. Hutter, and R. Siegwart, "Robust visual inertial odometry using a direct ekf-based approach," in *2015 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pp. 298–304, IEEE, 2015.

[34] P. Geneva, K. Eckenhoff, W. Lee, Y. Yang, and G. Huang, "Openvins: A research platform for visual-inertial estimation," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4666–4672, IEEE, 2020.

[35] L. Von Stumberg and D. Cremers, "Dm-vio: Delayed marginalization visual-inertial odometry," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 1408–1415, 2022.

[36] P. van Goor and R. Mahony, "Eqvio: An equivariant filter for visual-inertial odometry," *IEEE Transactions on Robotics*, 2023.