IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

FW-ORB-SLAM: A Monocular Visual SLAM Algorithm for Flapping-Wing Flying Robots

Zheng Zhong[1], Shou Chen[1], Qiang Fu[1], Jiubin Wang[1], Wei He[2], *Fellow, IEEE*

***Abstract*—Visual simultaneous localization and mapping (SLAM) is of great significance for flapping-wing flying robots (FWFRs) to enhance their autonomous navigation capabilities in complex environments. However, during the motion of FWFRs, there are intense image vibrations accompanied by significant illumination changes, which would prevent existing visual SLAM algorithms from being directly applied to FWFRs. Therefore, this paper proposes a modified ORB-SLAM3 algorithm called FW-ORB-SLAM for FWFRs. First, we adopt the fast Fourier transform (FFT) method to map the original images to the frequency domain. Then, based on the characteristic flapping motion of the FWFR, we decompose the frequency domain jitter to obtain stabilized images. Moreover, to mitigate the impact of illumination variations on feature point tracking during outdoor flight, a local adaptive contrast enhancement method is proposed, which enhances the stability of feature point tracking and augments the robustness of the SLAM algorithm. Finally, flight experiments carried out using our self-developed FWFR named U-Dove demonstrate that FW-ORB-SLAM outperforms the state-of-the-art ORB-SLAM3 algorithm, which provides insights into performing vision-based SLAM tasks for the FWFR.**

***Index Terms*—Biologically-Inspired Robots, SLAM.**

I. Introduction

FLAPPING-WING flying robots (FWFRs), representing a novel class of biomimetic robots, have attracted substantial attention in recent years owing to their exceptional maneuverability and stealth capabilities [1]. Due to the biomimetic design and flight patterns, FWFRs offer superior advantages in camouflage compared to rotary-wing and fixed-wing aerial vehicles. Equipped with visual sensors, FWFRs can be broadly deployed in missions including environmental reconnaissance, aerial surveying, and military reconnaissance. The successful execution of these tasks is heavily dependent on the autonomous navigation capabilities of the FWFR. Research on simultaneous localization and mapping (SLAM) constitutes a critical component in augmenting the autonomous navigation abilities of these robots [2].

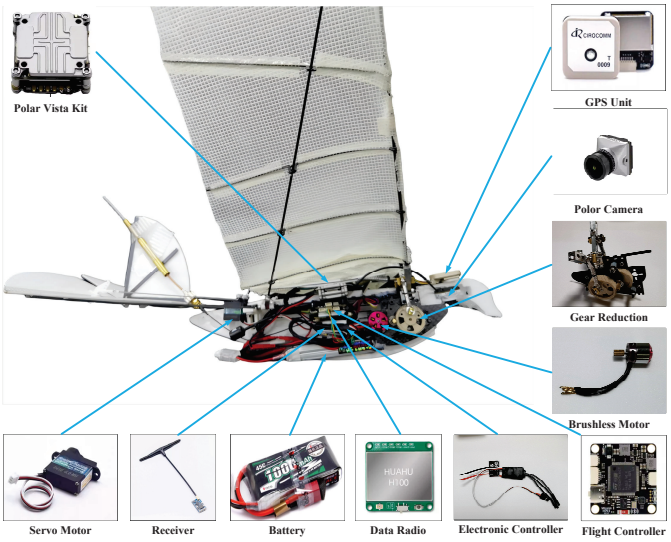

Fig. 1: Detailed component diagram of U-DOVE.

Visual SLAM is extensively employed in robot navigation and generally contains modules such as visual odometry, back-end optimization, and loop closure detection [3]. Certain state-of-the-art methods have exhibited remarkable performance, encompassing feature-based methods such as PTAM [4] and ORB-SLAM3 [5], direct methods like LSD-SLAM [6] and DSO [7], and multi-sensor fusion methods including VINS-Fusion [8] and Kimera [9]. However, owing to the flapping characteristics of the FWFR and the dynamic illumination variations in outdoor environments, existing visual SLAM methods are not directly applicable to FWFRs [10]. The GRIFFIN project [11] conducts research on outdoor visual SLAM for FWFRs by utilizing event cameras. But the event camera is costly and not well-suited for the routine tasks of the FWFR.

Nowadays, there are some studies on vision-based capabilities of the FWFR, such as target detection, target localization and obstacle avoidance. Nevertheless, research on outdoor visual SLAM for these robots remains a significant gap. The principal challenges contain high-frequency visual jitter due to flapping and illumination variations in dynamic outdoor environments. Furthermore, there is a scarcity of visual SLAM datasets for the FWFR, which are essential for validating the accuracy of different algorithms.

Manuscript received: May 15, 2025; revised August 12, 2025; Accepted October 25, 2025.

This paper was recommended for publication by Associate Editor Xinyu Liu upon evaluation of the Reviewers' comments. This work was supported in part by the National Natural Science Foundation of China under Grant 62173031, Grant 62225304, and Grant 62427813, in part by the Beijing Natural Science Foundation under Grant 25JL004, and in part by the Fundamental Research Funds for the China Central Universities under Grant FRF-TP-22-003C2. (Corresponding author: Qiang Fu)

[1]Zheng Zhong, Shuo Chen, Qiang Fu and Jiubin Wang are with the School of Intelligence Science and Technology and the Institute of Artificial Intelligence, University of Science and Technology Beijing, Beijing 100083, China, and also with the Key Laboratory of Intelligent Bionic Unmanned Systems, Ministry of Education, University of Science and Technology Beijing, Beijing 100083, China. (e-mail: D202310385@xs.ustb.edu.cn; m202410591@xs.ustb.edu.cn; fuqiang@ustb.edu.cn; b20200306@xs.ustb.edu.cn)

[2]Wei He is with School of Automation and Institute of Artificial Intelligence, Beijing Information Science and Technology University, Beijing 102206, China, and also with School of Intelligence Science and Technology and Institute of Artificial Intelligence, University of Science and Technology Beijing, Beijing 100083, China. (DOI): see top of this page.

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

To deal with the aforementioned challenges, we develope our own FWFR platform named U-Dove, as shown in Fig. 1. Based on this platform, we propose a modified ORB-SLAM3 algorithm called FW-ORB-SLAM for outdoor visual SLAM customized for FWFRs, thereby facilitating research on the autonomous navigation of FWFRs. The primary contributions of this study are as follows:

- We analyze the causes of image instability for FWFRs and introduce a frequency domain decomposition algorithm using the fast Fourier transform (FFT) to decompose high- and low-frequency jitter in FWFR images. Filters are then designed to achieve image stabilization.
- We design an adaptive image contrast enhancement and color correction algorithm to deal with the challenges posed by drastic outdoor illumination variations, thereby enhancing the accuracy of localization.
- Experiments are conducted on both the self-constructed dataset and the open-source FWAF-VID dataset [12]. These experiments verify the effectiveness of the proposed algorithm and provide directions for future research.

The remainder of this paper is structured as follows. Section II presents a concise overview of related research. A detailed description of our SLAM system is provided in Section III. Section IV introduces the comprehensive experimental configurations and presents the results obtained from our self-developed dataset. Section V introduces our experiments on the open-source FWAF-VID dataset, and discusses the advantages and limitations of this algorithm. Finally, Section VI provides a summary of our SLAM system and presents potential avenues for future research.

II. RELATED WORKS

A. Electronic Stabilization

Cameras offer advantages such as light weight, low power consumption, and rich image information, making them well-suited for deployment on FWFRs [13]. Many researchers have already conducted image stabilization studies for FWFRs, including both mechanical and electronic stabilization methods.

In the realm of mechanical stabilization, Fu et al. [14] draw inspiration from the double-concave structure of eagle eye to design a two-degree-of-freedom gimbal system with two cameras of different focal lengths. The dual cameras work collaboratively to balance wide-angle and telephoto perspectives dynamically, with the gimbal movements counteracting the high-frequency vibrations encountered during flight, thereby minimizing image blur and jitter. Pan et al. [15] develop a three-degree-of-freedom stabilizer specifically adapting to the flight dynamics of FWFRs, which addresses changes in posture and position. They also establish a dynamic and kinematic model of the stabilizer and devise a coordinated control strategy to ensure image stabilization.

Turning to electronic stabilization, Ye et al. [10] model the jitter resulting from the periodic flapping motion, and offer an alternative to traditional gimbal-based techniques. They introduce an adaptive trajectory adjustment algorithm based on jitter frequency, enhancing stabilization under complex motion. However, the algorithm robustness is compromised by uneven feature point distribution, leading to trajectory distortion. Liu et al. [16] leverage the distinct vibration patterns of the FWFR in flight, using the oriented fast and rotated brief (ORB) algorithm for feature extraction and matching to estimate motion. They estimate the flapping period from image features and design a dynamic sliding-window mean filter for motion filtering and stabilization, but the algorithm robustness in complex environments is limited. Yang et al. [17] optimize camera placement and image processing algorithms to mitigate the impact of vibrations on image quality and apply real-time image processing to counteract flight-induced vibrations, although the method's effectiveness is diminished under variable illumination conditions. Wang et al. [18] suggest a hybrid method that merges Kalman and low-pass filtering to address high-frequency jitter in video sequences from micro FWFRs, with optimized corner detection and optical flow computation for improved stabilization. Nevertheless, this approach is not well-suited for outdoor FWFR operations. Ye et al. [19] propose a binocular vision system based on long and short focus for image stabilization, which does not require GPU support and meets low-latency online requirements. At the same time, Ye et al. [20] also propose an image stabilization method of FWFRs based on IMU data, which solve the problem of optical flow tracking failure.

In summary, mechanical stabilization methods predominantly use gimbals, which unfortunately add to the payload and decrease the endurance of FWFRs. Electronic stabilization methods, while adding no payload, often do not fully consider the unique flight characteristics of FWFRs, resulting in complex algorithms that lack robustness in real applications. In this paper, we propose a practical frequency domain decomposition algorithm using FFT to decompose the high and low frequency jitter of FWFR image, and then design a filter to achieve electronic image stabilization.

B. Image Enhancement

Histogram equalization, a technique extensively employed in image enhancement, is particularly effective in enhancing image contrast [21]. This is achieved by redistributing the image brightness histogram to approximate a uniform distribution, thereby enhancing contrast. Techniques such as adaptive histogram equalization [22] and constrained contrast adaptive histogram equalization [23] are designed to preserve hue while enhancing contrast. Additionally, algorithms like brightness-preserving bi-histogram equalization [24] and binary sub-image histogram equalization methods have been developed. However, histogram equalization can lead to over-enhancement and other distortions. To address this, methods such as multi-scale Retinex with chromaticity preservation [25] and the use of multiple scale Gaussian filters have been proposed. Nevertheless, these methods often result in unrealistic phenomena such as over-enhancement and whitening in the enhanced images. To mitigate these issues, Fu et al. [26] propose a methodology that globally adjusts the initial estimated illumination and locally enhances contrast, which is then recombined with the reflectance to obtain the final

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

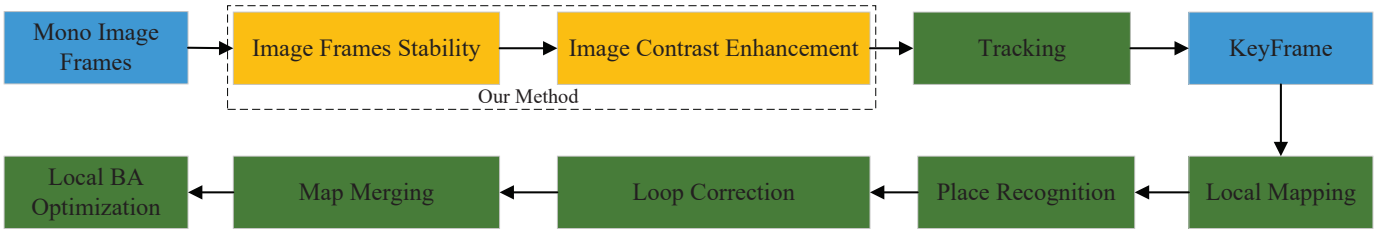

Fig. 2: Overall system architecture is outlined as follows. The system processes monocular images through image stabilization and local adaptive contrast enhancement, then performs image tracking and generates keyframes, and then performs local mapping. If it returns to the same area, it performs loopback detection and local optimization.

enhanced image. Wang et al. [27] introduce a multi-layer representation-based image contrast enhancement method that significantly improves robustness under dynamic illumination conditions but presents challenges in adapting to different images.

Deep learning methods are also commonly used in image enhancement. In the MSR-Net methodology, Shen et al. [28] manually curate high-quality data using photoshop software, followed by the application of random brightness adjustments, contrast reduction, and gamma transformation to produce enhanced images. However, this process entails manual preprocessing of the images. Zhang et al. [29] suggest incorporating hyperparameters and the Retinex model into the network architecture, which is a strategy that may potentially introduce noise. Jiang et al. [30] propose leveraging a generative adversarial network framework to learn the distribution of normally illuminated images, thereby constraining the brightness and contrast of the enhanced outcomes, although this method exhibits constraints in its adaptability to dynamic scenes. Luo et al. [31] propose a low-light image enhancement framework based on multi-view correlation, which performs well in noise suppression, detail retention and color consistency. Xie et al. [32] propose a method to restore low-light images to normal-light images through optimal global illumination distribution and clear local details, and achieve good results.

In the field of image enhancement, deep learning algorithms, while powerful, require substantial computational resources and are limited in their applicability. Traditional methods are often employed for overall image enhancement and may not focus on local details, which can result in over-enhancement of the image. In this paper, a local adaptive contrast enhancement algorithm is proposed to focus on the local details of the image to deal with the problem of illumination variations.

III. METHODOLOGY

A. System Overview

The comprehensive system architecture of FW-ORB-SLAM is illustrated in Fig. 2. The system input comprises monocular images acquired by the FWFR platform. These images undergo electronic stabilization and contrast enhancement processes prior to the extraction of ORB feature points, as shown in the orange part of Fig. 2. Image tracking is performed, and keyframes are generated. Subsequently, local mapping is conducted. When returning to the same area, loop detection and local optimization are carried out.

B. Video Stabilization Process

Firstly, the optical flow method is utilized to monitor the variations in optical flow between successive image frames. For two successive image frames, denoted as Frame I_1 and Frame I_2, the optical flow $\mathbf{u} = (u, v)$ indicates the displacement of each pixel over the time interval Δt. The position of each pixel is denoted by $\mathbf{x} = (x, y)$. The fundamental equation is presented in Eq. 1.

$$I_2(\mathbf{x}+\mathbf{u}, t) = I_1(\mathbf{x}, t) \tag{1}$$

We use the Farneback method [33] to get $\mathbf{u}$. Select a local window in the image, centered at pixel (x, y) and with a size of $(2n+1) \times (2m+1)$, as illustrated in Eq. 2:

$$E(\mathbf{u}) = \int_{x-n}^{x+n} \int_{y-m}^{y+m} \left(I(\mathbf{x}, t) - I(\mathbf{x}+\mathbf{u}, t+1)\right)^2 dxdy \tag{2}$$

Through the minimization of the aforementioned energy function $E(\mathbf{u})$, the optical flow vector $\mathbf{u}$ for each pixel is derived.

Subsequently, the optical flow magnitude, defined as the norm of the optical flow vector, is calculated to represent the displacement signal $d(t)$ in the image, as presented in Eq. 3.

$$d(t) = \sqrt{u^2 + v^2} \tag{3}$$

Then, the FFT transforms $d(t)$ into the frequency domain as $D(f)$:

$$D(f) = \mathscr{F}\{d(t)\} = \int_{-\infty}^{\infty} d(t) e^{-j2\pi ft} dt \tag{4}$$

where j signifies the imaginary unit, and f indicates the frequency.

Given that the t-th frame of the video contains n feature points, the pixel coordinates of the i-th point ($i \in [1, n]$) are $(x_{i,t}, y_{i,t})$. Similarly, its pixel coordinates in the $(t-1)$-th and $(t+1)$-th frames are $(x_{i,t-1}, y_{i,t-1})$ and $(x_{i,t+1}, y_{i,t+1})$, respectively. Due to the varying jitter acceleration of each feature point, $a(t)$ is computed as follows:

$$a(t) = \frac{1}{n} \sum_{i=1}^{n} \sqrt{\left(x_i^{t+1} + x_i^{t-1} - 2x_i^t\right)^2 + \left(y_i^{t+1} + y_i^{t-1} - 2y_i^t\right)^2} \tag{5}$$

Next, the flapping frequency can be estimated by peak detection and period segmentation based on $a(t)$ [10], and the corresponding flapping frequency is set as the cutoff frequency f_c. The FFT result $D(f)$ can be decomposed into

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

the high-frequency component $D_{\text{high}}(f)$ and the low-frequency component $D_{\text{low}}(f)$ as follows:

$$D_{\text{high}}(f) = \begin{cases} D(f) & |f| > f_c \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

$$D_{\text{low}}(f) = \begin{cases} D(f) & |f| \leq f_c \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

To transform the high-frequency and low-frequency components back to the time domain, we apply the inverse fast Fourier transform (IFFT). The IFFT of the high-frequency component $D_{\text{high}}(f)$ and the low-frequency component $D_{\text{low}}(f)$ can be expressed as follows:

$$\begin{cases} d_{\text{high}}(t) = \mathscr{F}^{-1}\{D_{\text{high}}(f)\} \\ d_{\text{low}}(t) = \mathscr{F}^{-1}\{D_{\text{low}}(f)\} \end{cases} \tag{8}$$

As for the separated high-frequency and low-frequency signals, the high-frequency signal $d_{\text{high}}(t)$ is the effect of motor vibration, which is removed by a low-pass filter. The low-frequency signal $d_{\text{low}}(t)$ is the effect of flapping, and a sliding window filter is designed based on the estimation of the flapping frequency to achieve image stabilization [16].

C. Image Contrast Enhancement

The local mean value is calculated as illustrated in Eq. 9:

$$u_B = \frac{1}{H * W} \sum_{i=x}^{x+W-1} \sum_{j=y}^{y+H-1} L_B(i,j) \tag{9}$$

where B denotes a local image block, u_B represents the average value of the image block, and L_B represents the corresponding grayscale matrix of B. Besides, (x,y) represents the central coordinate of local image block B; H and W denote the height and width of the image, respectively. Once the local mean value μ of the block B has been calculated, the local variance σ_B can be computed as follows:

$$\sigma_B = \frac{1}{H * W} \sum_{i=x}^{x+W-1} \sum_{j=y}^{y+H-1} L_B(i,j)^2 - u_B^2 \tag{10}$$

The average value of the local image patch is regarded as the low-frequency component. Subtracting this low-frequency component from the input image patch yields the high-frequency component. Additionally, an enhancement control factor is introduced to regulate the degree of enhancement of the high-frequency component, thereby preventing over-enhancement and over-saturation. Moreover, an enhancement cutoff factor is introduced to prevent excessive enhancement. The local contrast adjustment is carried out using Eq. 11, as detailed below:

$$\begin{cases} L_{EB}(i,j) = u_B + \alpha(L_B(i,j) - u_B), \alpha < \beta, \\ L_{EB}(i,j) = u_B + \beta(L_B(i,j) - u_B), \alpha \geq \beta. \end{cases} \tag{11}$$

where α represents the enhancement control factor, β is the enhancement cutoff factor, and L_{EB} denotes the enhanced local image block.

Next, guided filtering is employed to denoise and preserve edges in the enhanced local image block, as shown in Eq. 12.

$$L_{g_fE}(i,j) = \kappa(i,j) L_{N_E}(i,j) + \nu(i,j) \tag{12}$$

where $\kappa(i,j)$ and $\nu(i,j)$ are the linear coefficients within the local image block, and L_{N_E} is the guided image obtained through normalization.

The enhanced image is then color-corrected by converting it to the CIELAB color space. The retention of color channels is determined based on the pixel intensity of channels a and b. Color differences in other channels are compensated to make the image color appear more natural. Let I_c denote the average of channels a and b in the color space.

$$I_c = \frac{1}{H * W} \sum_{i=1}^{H} \sum_{j=1}^{W} I_c(i,j), c \in \{a,b\} \tag{13}$$

Based on the computed average color value, color balance compensation is applied to the other channels. After the color balance compensation, the balanced values for channels a and b are denoted as I_{ac} and I_{bc}, respectively.

$$\begin{cases} I_{ac} = I_a + \dfrac{\bar{I}_b - \bar{I}_a}{\bar{I}_b + \bar{I}_a} I_a, \bar{I}_a < \bar{I}_b. \\ I_{bc} = I_b + \dfrac{\bar{I}_a - \bar{I}_b}{\bar{I}_a + \bar{I}_b} I_b, \bar{I}_a > \bar{I}_b, \end{cases} \tag{14}$$

where I_a represents the pixel intensity of region a and $\bar{I}_a$ represents the average pixel intensity of region a. I_b represents the average pixel intensity of region b and $\bar{I}_b$ represents the average pixel intensity of region b. We present a flowchart of the image stabilization and local adaptive contrast enhancement, as illustrated in Fig. 3.

IV. EXPERIMENTAL RESULTS

The experiments are conducted on a Jetson ORIN NANO equipped with a 6-core CPU, running Ubuntu 20.04. The evaluation metric employed is the Absolute Trajectory Error (ATE) [12], measured in meters. The trajectory is aligned with the ground truth (i.e., onboard GPS data) utilizing SE(3) Umeyama alignment [5] prior to evaluation. To mitigate the randomness in the experimental outcomes, all the following results are derived from the average of ten repetitions. The parameters of U-Dove are shown in Table I.

TABLE I: Parameters of U-Dove.

Parameter	Value
Wingspan	80 cm
Body Weight	230 g
Flight Speed	5-10 m/s
Maximum Payload	100 g
Flapping Frequency	6-10 hz

A. Flapping-Wing Dataset

As shown in Fig. 1. U-Dove consists of GPS, gear reduction, brushless motor, flight controller, electronic speed controller, data radio, battery, receiver, servo motor, camera and image

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

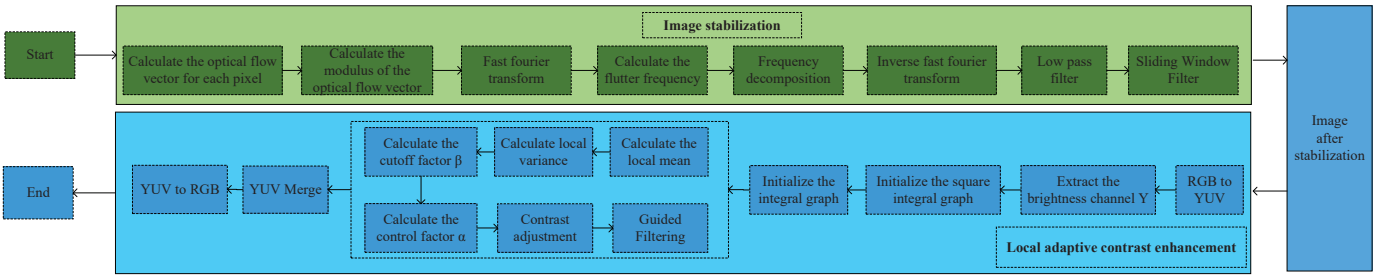

Fig. 3: Flowchart of the image stabilization and local adaptive contrast enhancement.

transmission module. The monocular dataset for FWFRs is collected using the platform U-Dove equipped with a vista camera. The ground truth trajectory values are recorded by the onboard GPS. To facilitate reference for interested researchers, some collected original images are provided, and the repository containing these original images is available at: https://github.com/asdfghjkl623/FW-ORB-SLAM.git

The dataset includes image data from four flight trajectories: straight (FW-01), slight turn (FW-02), sharp turn (FW-03), and circular (FW-04), as illustrated in Fig. 4. Table II summarizes the flight distance, flight duration, image resolution, and flight scenario of the flapping-wing dataset. All trajectories are collected outdoors, featuring an image resolution of 1920×1080 pixels. The flight distances range from 163 m to 461 m. The four trajectories of FW-01∼04 basically include the common flight modes of FWFR. The flight distances of the FW-01∼04 trajectories increase progressively, while their turning curvatures become increasingly larger.

TABLE II: Summary of the monocular dataset for FWFRs including aspects such as flight distance, flight duration, image resolution, and flight image scenario.

Trajectory	Length(m)	Time(s)	Resolution(pixel)	Scenario
FW-01	163.29	17.8	1920×1080	Outdoors
FW-02	255.26	29.8	1920×1080	Outdoors
FW-03	356.28	42.8	1920×1080	Outdoors
FW-04	460.35	59.8	1920×1080	Outdoors

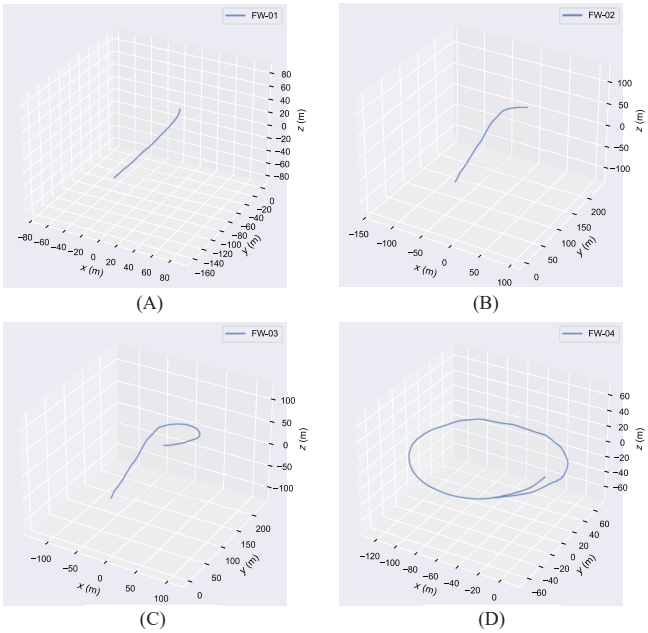

Fig. 4: Collecting the true value graph of the trajectories of the flapping-wing dataset. (A): Straight trajectory; (B): Slight turn trajectory; (C): Sharp turn trajectory; (D): Circular trajectory.

B. Video Stabilization Experiments

Image jitter for FWFRs is primarily caused by the coupling of wing flapping motion with motor vibration. As illustrated in Fig. 5, the jitter does not follow a regular pattern in the time domain. Real-time processing under conditions of turning or wind disturbance presents significant challenges.

Note that the wing flapping motion exhibits high amplitude and low frequency characteristics, whereas the motor vibration presents low amplitude and high frequency characteristics. We map the vibration from the time domain to the frequency domain using FFT, and separate the wing flapping motion from the motor vibration through frequency domain transformation and the application of a low-pass filter. Image jitter frequency caused by flapping-wing, estimated by Eq. 5, serves as the threshold for the low-pass filter. As depicted in Fig. 6, the red curve indicates the separated wing flapping motion, which exhibits clear upper-lower symmetry and conforms to the flapping pattern. The green curve represents the separated motor vibration, characterized by its high frequency nature.

After separating the high- and low-frequency jitter using FFT, we perform jitter compensation for each frequency band to obtain stabilized images. To validate the effectiveness of the jitter reduction algorithm, we assess the stabilization effect on the FW-01∼04 trajectories using Peak Signal-to-Noise Ratio (PSNR) [34] (in dB) as the metric. The larger the PSNR value, the higher the video stability. As shown in Table III, the deep-learning-based image stabilization method NNDVS [35] and the proposed image stabilization method are also evaluated. The proposed method consistently stabilizes all four test trajectories, with enhanced video stability observed for each. As shown in Table III, with PSNR as the evaluation index, the image stabilization effect of ORB-SLAM-stab is 11.74% higher than that of NNDVS, and 19.08% higher than that of ORB-SLAM3. At the same time, when using the NNDVS algorithm, it also takes up a lot of GPU resources, and the running time of the algorithm is significantly prolonged. Notably, ORB-SLAM3 integrated with the image stabilization module proposed herein is designated as ORB-SLAM3-Stab.

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

TABLE III: PSNR of the image stabilization effect tested on FW-01∼04.

Trajectory	ORB-SLAM3[dB]	NNDVS[dB]	ORB-SLAM3-Stab[dB]
FW-01	23.26	25.14	**27.13**
FW-02	22.76	23.97	**26.95**
FW-03	22.94	24.16	**27.25**
FW-04	21.87	23.52	**26.83**

C. Image Contrast Enhancement Experiments

During outdoor flight, FWFRs are subject to influences such as illumination change, which can result in uneven distribution or instability of feature points, thereby necessitating a considerable amount of time to track and update these points. Enhancing image contrast and achieving a more uniform color distribution can mitigate this issue. Fig. 7 (A) presents the original image, while Fig. 7 (B) shows the image after local adaptive contrast enhancement. It is evident that the enhanced image exhibits a more uniform color distribution, thereby facilitating feature tracking. The ORB-SLAM3-Stab algorithm with the addition of the proposed adaptive contrast enhancement module is named as FW-ORB-SLAM.

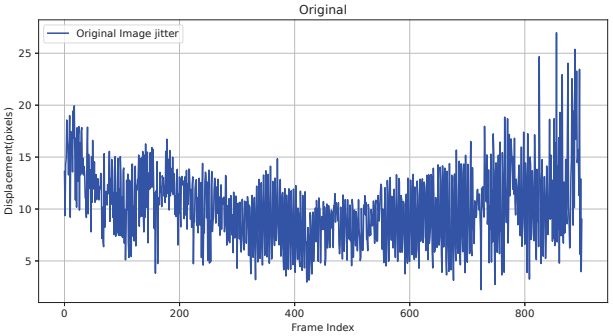

Fig. 5: Video jitter before image stabilization.

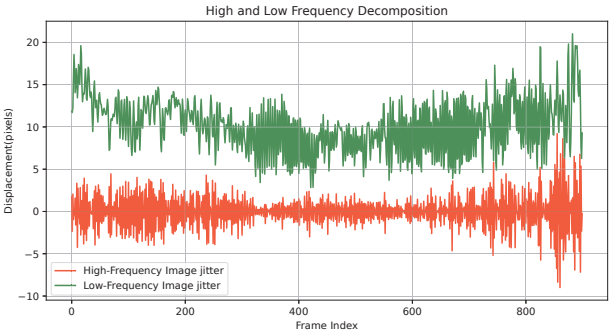

Fig. 6: Illustration of high-frequency jitter and low-frequency flapping after Fourier domain decomposition. Low-frequency flapping aligns with the FWFR's flapping pattern.

The tracking thread involves feature extraction, pose prediction, and keyframe selection. Analyzing its average runtime during FW-ORB-SLAM operation provides a more accurate assessment of the proposed image enhancement algorithm's effectiveness. As shown in Fig. 8, experiments on the FW-01∼04 trajectories reveal that compared with ORB-SLAM3-Stab, FW-ORB-SLAM with image enhancement significantly

Fig. 7: Comparison of images before and after local adaptive contrast enhancement. (A) represents the image before enhancement. (B) represents the image after enhancement.

reduces the tracking thread's average time consumption, highlighting the effectiveness of the proposed local adaptive contrast enhancement algorithm.

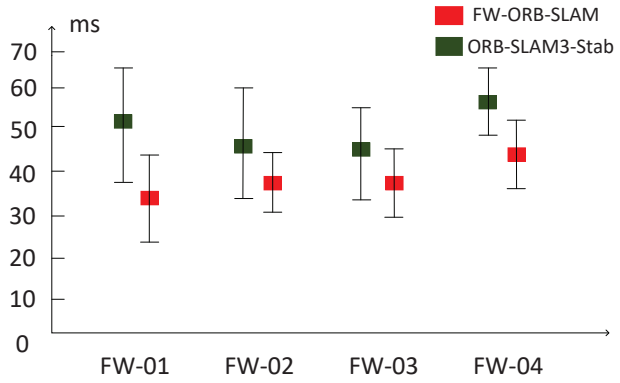

Fig. 8: Average consumption time of the tracking thread for FW-ORB-SLAM and ORB-SLAM3-Stab.

D. Comparative experiments

To assess the effectiveness of FW-ORB-SLAM, the ATE serves as the evaluation metric. The localization accuracy of the FW-ORB-SLAM algorithm is evaluated on the FW-01∼04 trajectories. As illustrated in Table IV, the original ORB-SLAM3 algorithm fails to operate successfully due to the strong jitter in FWFR images, and is marked as x. Although the ORB-SLAM3-Stab algorithm can run successfully after the addition of the proposed image stabilization module, its localization accuracy remains relatively low. Compared with ORB-SLAM3-Stab, FW-ORB-SLAM with the addition of the local adaptive contrast enhancement module enhances the localization accuracy by about 50%, because we reduce the influence of illumination variations on localization.

TABLE IV: Use ATE to evaluate the overall consistency of the trajectory for different SLAM algorithms.

Trajectory	ORB-SLAM3[m]	ORB-SLAM3-Stab[m]	FW-ORB-SLAM[m]
FW-01	x	0.0232	**0.0125**
FW-02	x	0.0713	**0.0397**
FW-03	x	0.2120	**0.1364**
FW-04	x	0.3606	**0.1567**

V. Discussion

To clarify the advantages and limitations of our algorithm, localization experiments are conducted on the FWAF-VID

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

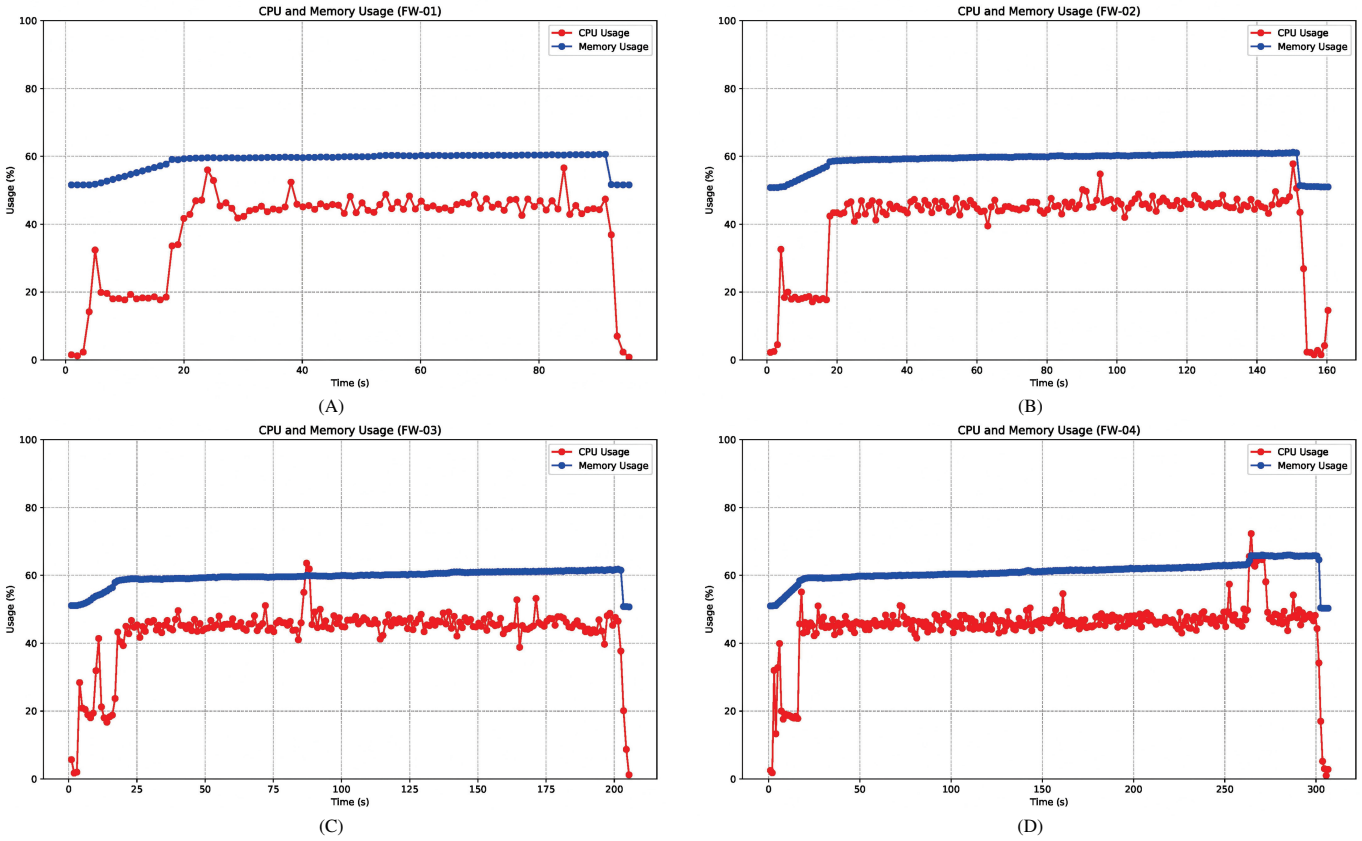

Fig. 9: The CPU and memory usage on Jetson, where red indicates CPU usage and blue indicates memory usage. (A) CPU and memory usage on FW-01. (B) CPU and memory usage on FW-02. (C) CPU and memory usage on FW-03. (D) CPU and memory usage on FW-04.

dataset [12], with Indoor16 and Indoor17 trajectories selected for testing. Indoor16 is an entirely dark indoor trajectory, while Indoor17 represents an indoor trajectory with alternating bright and dark conditions. Both trajectories involve continuous large-scale turns.

The ATE values of ORB-SLAM3 and FW-ORB-SLAM are tested under the Indoor16 and Indoor17 trajectories, as shown in Table V. Results indicate that FW-ORB-SLAM fails and cannot achieve localization in dark environments. In alternating bright and dark environments, it works, with positioning accuracy 6.8% higher than that of ORB-SLAM3. Overall, the positioning accuracy of FW-ORB-SLAM decreases significantly in alternating light and dark conditions and fails entirely in complete darkness. Robust positioning in dark environments will be addressed in future work.

TABLE V: Use ATE to evaluate the overall consistency of the trajectory for different SLAM algorithms.

Trajectory	ORB-SLAM3[m]	FW-ORB-SLAM[m]
Indoor16	x	x
Indoor17	14.41	**13.42**

Fig. 9 shows the real-time CPU and memory usage of FW-ORB-SLAM running on the FW dataset, indicating that it can operate in real time on the commonly-used Jetson airborne board. Table VI presents the processing frames per second (FPS) of FW-ORB-SLAM on the Jetson board, revealing a real-time running frame rate exceeding 12 FPS with CPU usage below 50%. Since the image size of the collected FW dataset is 1920×1080 while that of Indoor17 is 640×480, the FPS on Indoor17 is higher than that on the FW dataset as shown in Table VI. The system frame rate can be increased to 15 FPS by reducing the image size, enabling real-time processing requirements to be met during the general stable cruising of FWFRs. In general, the FW-ORB-SLAM algorithm proposed in this paper can satisfy the requirements for airborne real-time processing during stable in-air endurance, while the CPU and memory usage on the airborne board remains unsaturated. Future work will focus on improving the computational speed of the algorithm via parallel computing, thereby enhancing the real-time processing capability in terms of FPS.

TABLE VI: Running frames per second (FPS) of FW-ORB-SLAM algorithm on Jetson.

Trajectory	FW-01	FW-02	FW-03	FW-04	Indoor17
FW-ORB-SLAM	12.8	13.3	13.4	12.3	15.1

VI. CONCLUSIONS

In this paper, we propose a visual SLAM algorithm tailored for FWFRs on the basis of ORB-SLAM3. We find that the image jitter for FWFRs is primarily caused by the coupling of low-frequency, high-amplitude wing flapping motion with low-amplitude, high-frequency motor vibration. We transform this jitter from the time domain to the frequency domain for

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

decomposition, and subsequently perform electronic image stabilization based on the decomposition results. Moreover, to deal with illumination change during FWFR flight, we employ a local adaptive contrast enhancement method to enhance the stability of feature point tracking, thereby increasing the robustness of the algorithm. Extensive experiments have demonstrated the superior performance of the proposed visual SLAM algorithm. In the future work, we will continue to improve the robustness of the algorithm in challenging environments. At the same time, we will improve the FW-ORB-SLAM algorithm via parallel computing to improve the running speed of the algorithm, so as to improve the real-time processing of speed.

REFERENCES

[1] H. Huang, W. He, J. Wang, L. Zhang, and Q. Fu, "An all servo-driven bird-like flapping-wing aerial robot capable of autonomous flight," *IEEE/ASME Transactions on Mechatronics*, vol. 27, no. 6, pp. 5484–5494, 2022.

[2] J. Terblanche, S. Claassens, and D. Fourie, "Multimodal navigation-affordance matching for slam," *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 7728–7735, 2021.

[3] Y. Zhou, Y. Wang, F. Poiesi, Q. Qin, and Y. Wan, "Loop closure detection using local 3d deep descriptors," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 6335–6342, 2022.

[4] G. Klein and D. Murray, "Parallel tracking and mapping for small ar workspaces," in *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*. IEEE, 2007, pp. 225–234.

[5] 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.

[6] J. Engel, T. Schöps, and D. Cremers, "Lsd-slam: Large-scale direct monocular slam," in *European Conference on Computer Vision*. Springer, 2014, pp. 834–849.

[7] J. Engel, V. Koltun, and D. Cremers, "Direct sparse odometry," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 3, pp. 611–625, 2017.

[8] T. Qin, S. Cao, J. Pan, and S. Shen, "A general optimization-based framework for global pose estimation with multiple sensors," *arXiv preprint arXiv:1901.03642*, 2019.

[9] A. Rosinol, M. Abate, Y. Chang, and L. Carlone, "Kimera: an open-source library for real-time metric-semantic localization and mapping," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 1689–1696.

[10] J. Ye, E. Pan, and W. Xu, "Digital video stabilization method based on periodic jitters of airborne vision of large flapping wing robots," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 34, no. 4, pp. 2591–2603, 2023.

[11] J. P. Rodríguez-Gómez, R. Tapia, J. L. Paneque, P. Grau, A. G. Eguíluz, J. R. Martínez-de Dios, and A. Ollero, "The griffin perception dataset: Bridging the gap between flapping-wing flight and robotic perception," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 308–318, 2021.

[12] J. Jiang, E. Pan, W. Xu, W. Sun, and J. Ye, "Fwaf-vid: A flapping-wing aggressive flight benchmark dataset for visual-inertial localization," *IEEE Robotics and Automation Letters*, vol. 10, no. 6, pp. 6328–6335, 2025.

[13] Q. Fu, X. Chen, Z. Zheng, Q. Li, and W. He, "Research progress on visual perception system of bionic flapping-wing aerial vehicles," *Chinese Journal of Engineering*, vol. 41, no. 12, pp. 1512–1519, 2019.

[14] Q. Fu, S. Wang, J. Wang, X. Wu, and W. He, "Target detection and localization for flapping-wing robots based on biological eagle-eye vision," *IEEE/ASME Transactions on Mechatronics*, 2024, doi:10.1109/TMECH.2024.3424983.

[15] E. Pan, X. Liang, and W. Xu, "Development of vision stabilizing system for a large-scale flapping-wing robotic bird," *IEEE Sensors Journal*, vol. 20, no. 14, pp. 8017–8028, 2020.

[16] S. Liu, Q. Fu, N. Feng, C. Zhang, and W. He, "Design of an electronic image stabilization algorithm for flapping-wing flying robots," *Chinese Journal of Engineering*, vol. 46, no. 9, pp. 1544–1553, 2024.

[17] W. Yang, L. Wang, and B. Song, "Dove: A biomimetic flapping-wing micro air vehicle," *International Journal of Micro Air Vehicles*, vol. 10, no. 1, pp. 70–84, 2018.

[18] X. Wang, W. Zhang, J. Mou, and Z. Chen, "Design of real-time vision system applied in flapping-wing micro air vehicle," *Semiconductor Optoelectronics*, vol. 41, no. 1, pp. 114–117, 2020.

[19] J. Ye, E. Pan, J. Jiang, W. Sun, and W. Xu, "Gimbal-free online full-frame costabilization of long and short focus binocular vision for bionic flapping wing robots," *IEEE Transactions on Industrial Electronics*, pp. 1–12, 2025, doi:10.1109/TIE.2025.3585019.

[20] J. Ye, E. Pan, and W. Xu, "Real-time digital video stabilization based on imu data fusion and periodic jitters of airborne vision of bionic flapping wing robots," *IEEE Transactions on Instrumentation and Measurement*, vol. 74, pp. 1–16, 2025, doi:10.1109/TIM.2025.3552460.

[21] K. G. Dhal, A. Das, S. Ray, J. Gálvez, and S. Das, "Histogram equalization variants as optimization problems: a review," *Archives of Computational Methods in Engineering*, vol. 28, pp. 1471–1496, 2021.

[22] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J. B. Zimmerman, and K. Zuiderveld, "Adaptive histogram equalization and its variations," *Computer Vision, Graphics, and Image Processing*, vol. 39, no. 3, pp. 355–368, 1987.

[23] E. D. Pisano, S. Zong, B. M. Hemminger, M. DeLuca, R. E. Johnston, K. Muller, M. P. Braeuning, and S. M. Pizer, "Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms," *Journal of Digital Imaging*, vol. 11, pp. 193–200, 1998.

[24] Y. Wang, Q. Chen, and B. Zhang, "Image enhancement based on equal area dualistic sub-image histogram equalization method," *IEEE Transactions on Consumer Electronics*, vol. 45, no. 1, pp. 68–75, 1999.

[25] D. J. Jobson, Z.-u. Rahman, and G. A. Woodell, "A multiscale retinex for bridging the gap between color images and the human observation of scenes," *IEEE Transactions on Image Processing*, vol. 6, no. 7, pp. 965–976, 1997.

[26] X. Fu, D. Zeng, Y. Huang, Y. Liao, X. Ding, and J. Paisley, "A fusion-based enhancing method for weakly illuminated images," *Signal Processing*, vol. 129, pp. 82–96, 2016.

[27] X. Wang, M. Christie, and E. Marchand, "Optimized contrast enhancements to improve robustness of visual tracking in a slam relocalisation context," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 103–108.

[28] L. Shen, Z. Yue, F. Feng, Q. Chen, S. Liu, and J. Ma, "Msr-net: Low-light image enhancement using deep convolutional network," *arXiv preprint arXiv:1711.02488*, 2017.

[29] Y. Zhang, J. Zhang, and X. Guo, "Kindling the darkness: A practical low-light image enhancer," in *Proceedings of the 27th ACM International Conference on Multimedia*, 2019, pp. 1632–1640.

[30] Y. Jiang, X. Gong, D. Liu, Y. Cheng, C. Fang, X. Shen, J. Yang, P. Zhou, and Z. Wang, "Enlightengan: Deep light enhancement without paired supervision," *IEEE Transactions on Image Processing*, vol. 30, pp. 2340–2349, 2021.

[31] H. Luo, B. Chen, L. Zhu, P. Chen, and S. Wang, "Rcnet: Deep recurrent collaborative network for multi-view low-light image enhancement," *IEEE Transactions on Multimedia*, vol. 27, pp. 2001–2014, 2025.

[32] Z. Xie, H. Ren, J. Huang, Z. He, H. Lu, Y. Liu, J. Lu, L. Yuan, S. Liu, and C. Xie, "Low-light image enhancement via multi-exposure progressive contrastive regularization," *IEEE Transactions on Circuits and Systems for Video Technology*, 2025, doi:10.1109/TCSVT.2025.3589217.

[33] G. Farnebäck, "Two-frame motion estimation based on polynomial expansion," in *Image Analysis: 13th Scandinavian Conference, SCIA 2003 Halmstad, Sweden, June 29–July 2, 2003 Proceedings 13*. Springer, 2003, pp. 363–370.

[34] A. Hore and D. Ziou, "Image quality metrics: Psnr vs. ssim," in *2010 20th International Conference on Pattern Recognition*. IEEE, 2010, pp. 2366–2369.

[35] Z. Zhang, Z. Liu, P. Tan, B. Zeng, and S. Liu, "Minimum latency deep online video stabilization," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 23 030–23 039.