

ROBUST MOTION ESTIMATION FOR INDOOR MAVS USING WIDE-FIELD-INTEGRATION OF OPTIC FLOW

Masaya Yanagisawa[1], Ryuichiro Okabayashi[1], Tomoka Yamaguchi[2], Mitsuki Takamoto [2], Mai Bando[1] & Shinji Hokamoto[1]

[1] Department of Aerospace Engineering, Kyushu University
744 Motooka, Nishi-ku, Fukuoka City, Fukuoka 819-0395, JAPAN
TEL: +81-92-802-0006
yanagisawa.masaya.769@s.kyushu-u.ac.jp, okabayashi.ryuichiro.176@s.kyushu-u.ac.jp
mbando@aero.kyushu-u.ac.jp, hokamoto@aero.kyushu-u.ac.jp
[2] Advanced Technology R&D Center, Mitsubishi Electric Corporation
8-1-1, Tsukaguchi-Honmachi, Amagasaki City, Hyogo 661-8661, JAPAN
TEL: +81-6-6497-7699
Yamaguchi.Tomoka@dp.MitsubishiElectric.co.jp, Takamoto.Mitsuki@aj.MitsubishiElectric.co.jp

Abstract

This paper proposes a robust motion estimation framework for Micro Air Vehicles (MAVs) based on Wide-Field-Integration (WFI) of optic flow, which is useful for indoor flight of MAVs. In GNSS denied environments, optic flow obtained from image sensors is degraded by several factors: mismatches in feature-point tracking, the presence of different depth surfaces, and featureless regions. To prevent estimation degradation due to these effects, this study proposes a method to isolate unreliable optic flow and utilize only reliable optic flow vectors for WFI of optic flow. The effectiveness of the proposed method is verified through computer simulations and also preliminary experiments using a linear slider system. The experimental results show that the proposed method can reduce the root mean square error of translational motion estimation by more than 80%.

Keywords: Wide-Field-Integration of Optic Flow, Outlier Rejection, Motion Estimation, Autonomous Navigation, Micro Air Vehicles

1. Introduction

In recent years, the utilization of Micro Air Vehicles (MAVs) has expanded into indoor and underground environments where Global Navigation Satellite System (GNSS) signals are unavailable [1]. In such GNSS-denied environments, autonomous navigation relies on typical onboard sensors such as LiDAR or stereo vision systems. However, these conventional approaches often involve significant weight and computational penalties. For instance, while Time-of-Flight LiDAR systems enable active and dense 3D mapping, their inherent power consumption and sensor weight impose significant constraints on the design and flight endurance of ultra-lightweight MAVs [2]. Visual Simultaneous Localization and Mapping (vSLAM) or stereo vision is a widely adopted alternative; however, typical algorithms require intensive computational resources for depth estimation and a sufficient baseline width of cameras, which is often physically constrained on small platforms [3]. Furthermore, vision-based localization frequently fails in environments with low-texture features, such as white corridors, leading to a loss of estimation stability [4]. In fact, recent trends in aerial robotics emphasize the need for frugal, lightweight visual navigation strategies that can operate reliably even with minimal computational resources and narrow fields of view [5].

To address these challenges, our research group has focused on the Wide-Field-Integration (WFI) of optic flow as a bio-inspired, computationally efficient alternative [6, 7, 8]. Inspired by the visual processing system of flying insects, the WFI scheme estimates motion variables through a direct linear

integration of flow vectors over a wide field of view [9]. Unlike conventional vSLAM or stereo-matching algorithms relying on heavy iterative convergence calculations or complex geometric environmental information, the WFI of optic flow enables low-latency velocity and attitude feedback without explicit 3D map generation. The fundamental effectiveness and stability of this WFI approach for autonomous flight control have been extensively demonstrated in previous foundational studies [10, 11].

Despite these architectural improvements, conventional WFI of optic flow fundamentally relies on optic flow acquisited from a flat and feature-rich plane. In practical indoor scenarios, this assumption is frequently violated, and its estimation accuracy can be degraded by three primary factors: (1) inherent errors or noise in the flow generation algorithm itself; (2) the intrusion of obstacles or walls into the camera's field of view, which changes the distance to a virtually defined planar surface; and (3) the inability to detect feature points in low-texture regions. These factors are particularly critical in indoor environments, which are often happened in dim lighting, narrow passages with close obstacles, and texture-less surfaces. Under such conditions, "outliers"—flow vectors deviating significantly from ideal ones—are frequently generated. Since the WFI algorithm utilizes all acquired flow vectors in its integration process, these outliers can severely bias and fluctuate the motion estimation results. This significant loss of estimation accuracy directly compromises the vehicle's control, ultimately leading to unstable flight or collisions with surrounding obstacles.

The primary objective of this study is to develop a highly robust motion estimation system capable of maintaining stable estimation even when the WFI algorithm is exposed to severe visual degradation in indoor environments. We propose a robust motion estimation method that improves conventional WFI of optic flow by identifying and rejecting outliers. The proposed framework integrates spatial masking with a motion prediction model based on recent kinematic motion. By leveraging past motion states to examine current flow data, the system can safely bridge periods of extreme visual noise—such as crossing glossy or low-texture surfaces—without experiencing computational divergence. The effectiveness of the proposed architecture is validated through both numerical simulations and real-world preliminary experiments under intentionally degraded visual conditions.

2. Motion-Variable Estimation through Wide-Field-Integration (WFI) of Optic Flow

2.1 Estimation Process

Optic flow is the vector field of relative velocities between an image sensor and its surrounding environment. Consider a spherical image plane of unit radius centered at the image sensor placed at a vehicle's center of mass. The geometric relationship between the coordinate frame and the target environment is illustrated in Fig. 1. A point on the spherical surface can then be described by its azimuth angle γ and depression angle β as

$$Q = [\sin\beta\cos\gamma \;\; \sin\beta\sin\gamma \;\; \cos\beta]^T \tag{1}$$

When the vehicle has translational velocity $v = [u\, v\, w]^T$ and angular velocity $\omega = [p\, q\, r]^T$, the optic flow $\dot{Q}$ measured on the spherical surface is expressed as(see [8])

$$\dot{Q} = -\omega \times Q - \mu\{v - (v^T Q)Q\} \tag{2}$$

where μ is the nearness function defined by the inverse of the distance to a feature point of the environment. The WFI of optic flow process virtually assumes a flat surface, and thus the nearness function is expressed as follows:

$$\mu = \frac{\sin\beta\cos\gamma\sin\theta - \sin\beta\sin\gamma\sin\phi\cos\theta - \cos\beta\cos\phi\cos\theta}{z} \tag{3}$$

where z is the altitude of the vehicle and ϕ and θ are its roll and pitch angles defined in the 3-2-1 Euler angles.

To reduce the effect of flow-noise or unevenness of the surface on motion-variable estimation, optic flow is integrated over wide ranges by using sensitivity functions in the WFI of optic flow process. Spherical harmonics are used for the sensitivity functions, and for a practical computational cost, the continuous integral can be replaced by a Riemann sum without degrading estimation accuracy

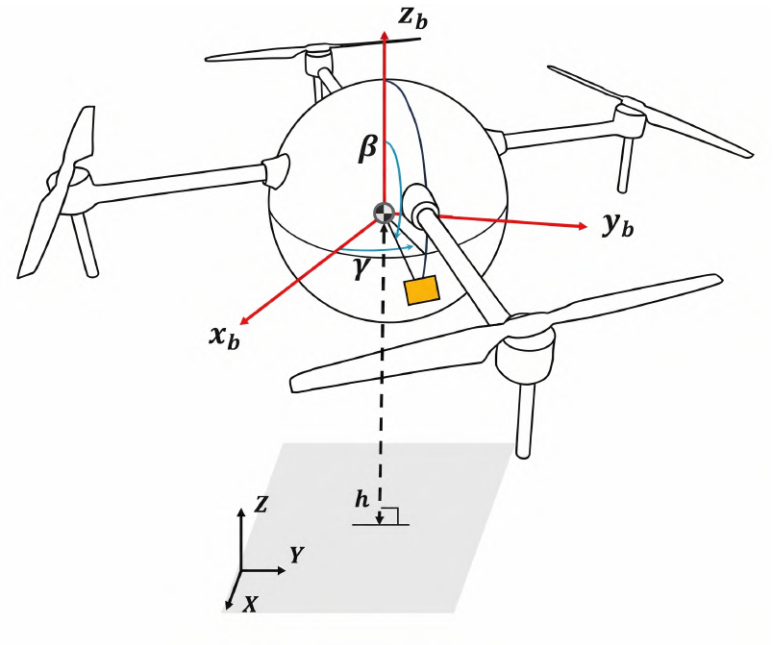


Figure 1 – Coordinate system and geometric relationship of the optic flow observation.

[6]. Furthermore, to preserve the independence of the two orthogonal components of the optic flow, the sensitivity functions are defined separately for each direction. Consequently, the output vector y, obtained from the flow at discrete points (total number: N), can be expressed using spherical harmonic functions up to order M (F_1-F_M) as follows:

$$y = \sum_{k=1}^{N} \begin{bmatrix} \dot{Q}^{\gamma}(\gamma_k,\beta_k)F_1^{\gamma}(\gamma_k,\beta_k) \\ \vdots \\ \dot{Q}^{\gamma}(\gamma_k,\beta_k)F_M^{\gamma}(\gamma_k,\beta_k) \\ \dot{Q}^{\beta}(\gamma_k,\beta_k)F_1^{\beta}(\gamma_k,\beta_k) \\ \vdots \\ \dot{Q}^{\beta}(\gamma_k,\beta_k)F_M^{\beta}(\gamma_k,\beta_k) \end{bmatrix} \mathrm{d}\Omega_k \tag{4}$$

where the superscripts γ and β denote the directions of the flow, and $\mathrm{d}\Omega_k$ represents the solid angle of the k-th pixel.

On the other hand, the output y can also be expressed based on the virtual plane assumption. This formulation utilizes the linear relationship between the optic flow and the motion variables $x = [u\ v\ w\ p\ q\ r]^T$, as shown in Eq. (2). Thus, the integration can be equivalently expressed as:

$$y = Cx \tag{5}$$

where C is the $2M \times 6$ coefficient matrix. Specifically, this matrix is explicitly expanded as follows:

$$C = \sum_{k=1}^{N} \begin{bmatrix} \mu_k \sin\gamma_k F_1^\gamma & -\mu_k \cos\gamma_k F_1^\gamma & 0 \\ \vdots & \vdots & \vdots \\ \mu_k \sin\gamma_k F_M^\gamma & -\mu_k \cos\gamma_k F_M^\gamma & 0 \\ -\mu_k \cos\beta_k \cos\gamma_k F_1^\beta & -\mu_k \cos\beta_k \sin\gamma_k F_1^\beta & \mu_k \sin\beta_k F_1^\beta \\ \vdots & \vdots & \vdots \\ -\mu_k \cos\beta_k \cos\gamma_k F_M^\beta & -\mu_k \cos\beta_k \sin\gamma_k F_M^\beta & \mu_k \sin\beta_k F_M^\beta \end{bmatrix} \begin{bmatrix} \cos\beta_k \cos\gamma_k F_1^\gamma & \cos\beta_k \sin\gamma_k F_1^\gamma & -\sin\beta_k F_1^\gamma \\ \vdots & \vdots & \vdots \\ \cos\beta_k \cos\gamma_k F_M^\gamma & \cos\beta_k \sin\gamma_k F_M^\gamma & -\sin\beta_k F_M^\gamma \\ \sin\gamma_k F_1^\beta & -\cos\gamma_k F_1^\beta & 0 \\ \vdots & \vdots & \vdots \\ \sin\gamma_k F_M^\beta & -\cos\gamma_k F_M^\beta & 0 \end{bmatrix} \mathrm{d}\Omega_k \tag{6}$$

By using the pseudo-inverse of the matrix C, the motion variables x can be estimated as follows:

$$x = C^\dagger y = (C^T C)^{-1} y \tag{7}$$

When applying the above estimation method to a real system, it is necessary to bridge the gap between the assumptions in the estimation theory and the actual system. First, while the theory considers optic flow on a spherical image plane, standard image sensors obtain optic flow on planar image planes. Thus, the measured optic flow must be transformed onto a spherical image representation. In addition, the field of view(FOV) of a single image sensor is limited, and thus different motion variables may show similar flow patterns in the limited FOV, creating estimation ambiguity. To address this issue, it is effective to use multiple image sensors and integrate optic flow obtained from spatially separated regions. Furthermore, it is generally difficult to place image sensors exactly at the vehicle's center of mass, and they are typically mounted at locations offset from it. In such cases, a transformation matrix D is introduced to relate the motion at the sensor locations to that at the vehicle's center of mass. By incorporating these procedures as detailed in previous studies [9, 12], the output vectors and matrices from n cameras are vertically stacked, and as a result the motion variables of the vehicle can be estimated using the following equation:

$$x = \begin{bmatrix} C_1 D_1 \\ C_2 D_2 \\ \vdots \\ C_n D_n \end{bmatrix}^\dagger \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \tag{8}$$

2.2 Excluding Process of Unreliable Optic Flow

As described above, the motion variables of a drone can be estimated by first computing the output vector y from all optic flow measured by the image sensor, and then multiplying the pseudo-inverse of the matrix C from the left to obtain x as shown in Eq.(7). In this process, errors in optic flow caused by misdetection of feature points due to illumination conditions or by small unevenness of the observation terrain can be mitigated by the integration process of optic flow over a wide region. However, when obstacles or walls appear in the image, the virtual plane defined as an averaged terrain may be formed differently from the actual plane. As a result, measuring a distance to the virtual plane becomes difficult even by utilizing an altitude sensor. In addition, when featureless regions exist on the terrain, detecting feature points in such areas becomes unreliable. In this case, the optic flow obtained by tracking feature points is likely to connect unrelated points, resulting in a vector field with inconsistent magnitudes and directions. These unreliable optic flow vectors lead to canceling each other out in the integration process of Eq.(6), and the magnitude of the summation in Eq.(4) may be reduced considering the number of optic flow vectors.

To address this issue, this study excludes optic flow vectors derived from surfaces other than the assumed virtual plane, as well as unreliable flow vectors extracted from featureless regions or noises. By estimating the motion variables using only reliable optic flow measurements, the estimation accuracy is improved. In the formulation, this extraction of unreliable flow corresponds to performing the summations in Eq. (4) and in the matrix C using only reliable optic flow vectors. It should be noted that, as long as the total number of optic flow vectors N does not become excessively small, the degradation in estimation accuracy remains limited.

The reliability of optic flow obtained at an instant can be assessed by comparing it with the flow predicted by extrapolating the drone's motion variables from the recent motion. If the drone continues approximately the same motion, its translational and angular velocities are basically the same as previously estimated values. The attitude is obtained by integrating the angular velocity, and the altitude is estimated from the velocity and attitude. Based on these quantities, the nearness function μ with respect to the virtual plane is computed from Eq. (3), and the corresponding optic flow can be calculated from Eq. (2) using the nearness function together with the translational and angular velocities. Even if the drone changes its motion, as long as the motion change can be reasonably predicted, the optic flow observed by the onboard sensors would be estimated in the same manner. Therefore, by comparing the predicted optic flow $\hat{\dot{Q}}$ with the observed raw one $\dot{Q}_{raw}$, not only random flow noise but also flow vectors obtained from surfaces other than the virtual plane and those obtained from unreliable matching in featureless regions can be identified as outliers. This identification is conducted by introducing an error threshold T_{th} for the flow magnitude, generating a binary mask M_k for each spatial data point k:

$$M_k = \begin{cases} 1 & \text{if } |\dot{Q}_{raw}^{(k)} - \hat{\dot{Q}}^{(k)}| < T_{th} \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

Data points where $M_k = 0$ are effectively excluded from the subsequent spatial integration. It should be noted that the computational cost remains low, because the outlier rejection processes rely on extrapolation, basic arithmetic operations, and simple thresholding.

As a practical example of the extrapolation process, a discrete median-based kinematic extrapolation is employed in this study, as illustrated in Fig. 2. To robustly handle severe noise and arbitrary flow vectors, a history buffer maintains the recent valid estimates. Assuming a nearly constant sampling interval Δt between consecutive image frames, the discrete state difference for the motion variables Δx is evaluated. The representative variation step is determined by taking the median of the differences between the recent valid states. This approach inherently prevents momentary severe outliers in the history buffer from skewing the prediction. Let $\alpha \in (0, 1]$ be a relaxation parameter applied to scale the trend, preventing over-extrapolation caused by natural motion fluctuations. The reference state vector at the current step t is predicted as:

$$\hat{x}(t) = x(t-1) + \Delta x \cdot \alpha \tag{10}$$

By substituting this predicted motion $\hat{x} = [\hat{v}^T \ \hat{\omega}^T]$ into Eq. (2), the theoretical optic flow $\hat{\dot{Q}}$ can be calculated to perform the robust outlier rejection seamlessly.

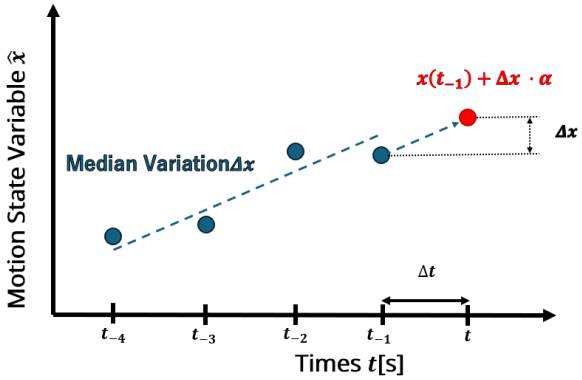


Figure 2 – Concept of discrete median-based kinematic extrapolation for motion prediction.

3. Verification of Proposed Method in Numerical Simulation

3.1 Simulation Setup and Error Modeling

To verify the effectiveness of the proposed method, numerical simulations are conducted by using a model of MAV flying above a flat terrain. The simulation conditions and camera configuration are summarized in Table 1.

Table 1 – Simulation parameters and sensor specifications

Parameter	Symbol	Unit	Value
Vehicle CoG altitude	z	[m]	1.0
Translational velocity	(u, v, w)	[m/s]	$(1.0, 1.0, 0.0)$
Angular velocity	(p, q, r)	[rad/s]	$(0.0, 0.0, 0.0)$
Cam 1 mounting config.	$(r_{c1}, \gamma_{c1}, \beta_{c1})$	[m, deg, deg]	$(0.05, 0, 135)$
Cam 1 optical axis	$(\gamma_{op,1}, \beta_{op,1})$	[deg, deg]	$(0, 135)$
Cam 2 mounting config.	$(r_{c2}, \gamma_{c2}, \beta_{c2})$	[m, deg, deg]	$(0.05, 90, 135)$
Cam 2 optical axis	$(\gamma_{op,2}, \beta_{op,2})$	[deg, deg]	$(90, 135)$
Field of View (H $\times$ V)	FOV	[deg]	95×70
Image resolution (H $\times$ V)	-	[pixels]	600×400
Flow grid resolution	$N_x \times N_y$	[points]	30×20
Sampling rate	f_s	[Hz]	15
Total flight duration	T	[s]	20

The MAV is assumed to maintain its altitude of $z = 1.0$ m keeping a level attitude ($\phi = 0, \theta = 0$). Its horizontal velocity components along the body's x and y directions are both 1.0 m/s ($u = v = 1.0$ m/s), while the vertical velocity is set to zero. Two cameras are assumed to be mounted at positions located 0.05 m from the center of mass, with azimuth/elevation angles of ($0°$, $135°$) and ($90°$, $135°$), respectively, and with their optical axes oriented in the same directions. Each camera has a field of view of $95°$ in the horizontal direction and $70°$ in the vertical direction, and assumes an image resolution of 600×400 pixels. From this pixel data, 30×20 optic flow vectors are measured in its flat image plane. These measurements are then mapped onto a spherical image surface to apply the WFI of optic flow algorithm. The flight time of the simulation is 20 s and its sampling rate of the optic flow is 15 Hz.

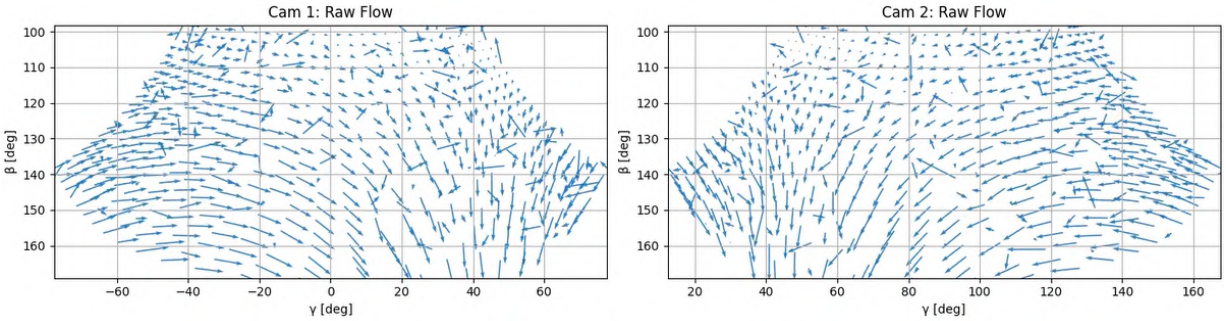
On the other hand, to emulate optic flow under practical conditions, noises in flow extraction processes, flow originating from surfaces different from the virtual plane, and errors due to featureless regions are taken into account. Specifically, Gaussian noise with zero mean and a variance of 0.05 rad/s is first added to the optic flow generated from the drone's true motion. This noise is set based on the characteristics of the imaging system identified in our experiments. Moreover, out of the 600 flow vectors obtained from a single camera, 180 are randomly selected, and their magnitudes are modified by multiplying uniformly distributed random values in the range $[0, 1]$. In addition, Gaussian noise with zero mean and a variance of 0.50 rad/s is then added to these selected flow vectors. Such errors are added to the true optic flow sampled at 15 Hz over 20 s, and the resulting flow is regarded as image sensor measurements in practical situations. The WFI of optic flow estimation procedure is applied to these data.

3.2 Visual Analysis of Outlier Rejection Performance

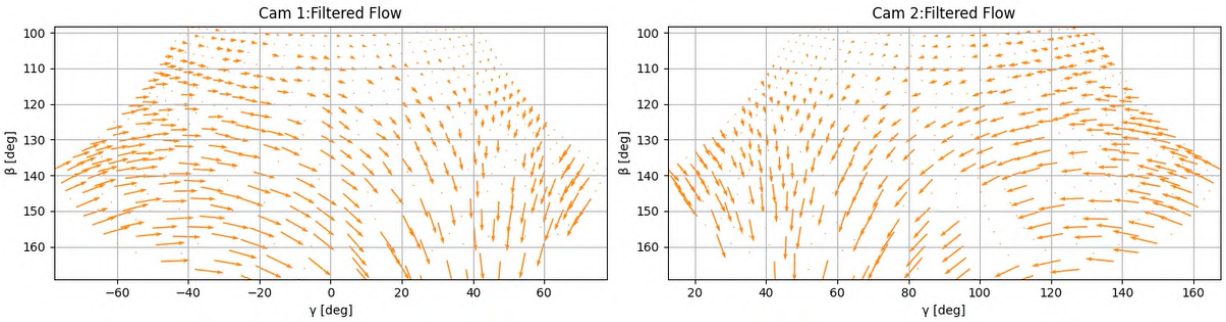
The performance of the outlier elimination is inspected visually in the optic flow. Figure 3a shows an example of realistic optic flow obtained at a given instant from two cameras under the above conditions. As illustrated in the figure, due to the aforementioned disturbances, some flow vectors have different magnitudes and directions from their surrounding ones. These corrupted vectors significantly deviate from the ideal flow generated by the actual motion, making it difficult for the standard WFI of optic flow method to accurately estimate the vehicle's motion.

Figure 3b shows the filtered flow vectors after applying the proposed rejection process. As seen in the figure, the highly inconsistent flow vectors are successfully eliminated, leaving only the reliable

flow vectors that represent the correct motion state. This cleaned input data establishes a reliable foundation for the subsequent matrix reduction phase and stable state estimation.



(a) Raw corrupted flow field (30% random outliers)



(b) Filtered flow field after applying the proposed spatial mask

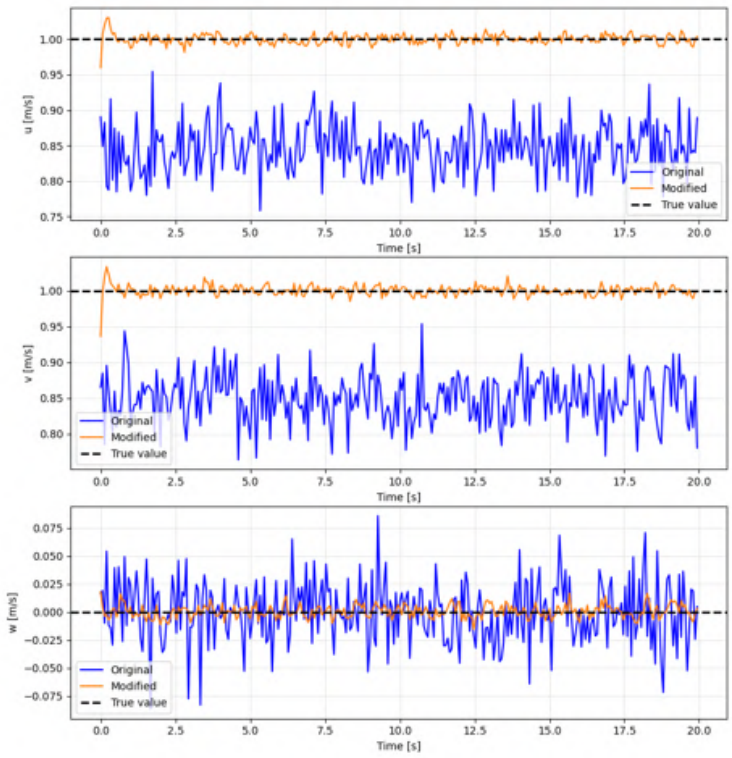
Figure 3 – Visual comparison of integrated optic flow fields for the dual-camera system. In the raw field (a), 30% of the vectors are severely corrupted by arbitrary noise, obscuring the underlying kinematic pattern. After applying the spatial mask (b), gross errors are effectively rejected, and only the valid vectors consistent with the reference state are retained for the matrix reduction.

3.3 Evaluation of Motion Estimation Accuracy

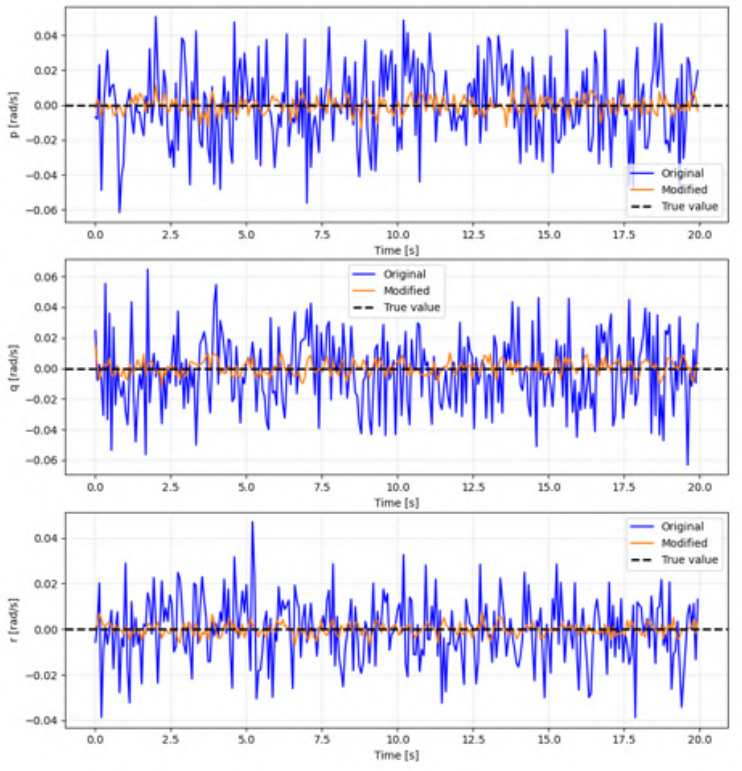
To evaluate the overall performance of the motion estimation framework, the motion variables estimated from both the standard WFI of optic flow (Original method) and the proposed method are compared with the true motion.

Figure 4 shows the time-series comparison of the estimated motion variables under the 30% outlier condition. As observed in the figure, the Original method suffers from severe, high-frequency oscillations and large deviations from the true values. These errors are a direct consequence of 30% outliers.

In contrast, the proposed method demonstrates accurate tracking performance across all motion variables. The estimation results of the proposed framework stably average around the ground truth, exhibiting a significantly smaller standard deviation compared to the baseline approach. The results indicate that the system can maintain high numerical stability and minimal variance even in the presence of extreme visual corruption.



(a) Comparison of translational velocities (u, v, w)



(b) Comparison of angular velocities (p, q, r)

Figure 4 – Time-series comparison of the estimated motion variables under 30% random outlier noise. The plots clearly demonstrate the transient stability of the Proposed method (orange solid line) against the significant spikes of the Original method (blue solid line) compared to the ground truth (black dashed line).

To provide a quantitative assessment, the Root Mean Square Error (RMSE) is calculated for the entire duration of the simulation ($T = 20$ s). The estimation performance is further evaluated using the reduction rate R [%], defined as:

$$R = \left(1 - \frac{RMSE_{proposed}}{RMSE_{original}}\right) \times 100 \tag{11}$$

where $RMSE_{original}$ and $RMSE_{proposed}$ denote the errors of the baseline and the proposed framework, respectively. The results are summarized in Table 2.

Table 2 – Quantitative comparison of estimation errors (RMSE)

Variable	Original RMSE	Proposed RMSE	Reduction Rate [%]
u [m/s]	0.158	0.007	95.5
v [m/s]	0.154	0.007	95.2
w [m/s]	0.027	0.006	79.0
p [rad/s]	0.022	0.004	80.8
q [rad/s]	0.024	0.005	80.6
r [rad/s]	0.014	0.002	82.5

As shown in Table 2, the Proposed method achieves a substantial reduction in estimation error across all motion variables. Notably, the translational velocities in the horizontal plane (u, v) exhibit a reduction rate exceeding 95%, which quantitatively confirms the significant decrease in variance and large deviations from the true values observed in the time-series analysis. Even for the rotational components, the proposed framework achieves an average improvement of over 80%. These results objectively demonstrate that the proposed method maintains high estimation accuracy in severely degraded visual environments.

4. Preliminary Experimental Results

Following the numerical evaluations in Section 3, this section presents the experimental validation of the proposed framework using a hardware system. In this study, a linear slider system is employed instead of a free-flying MAV. This setup isolates the motion variables to a specific axis, eliminating unpredictable aerodynamic disturbances and ensuring that the accuracy of the motion estimation can be rigorously compared against high-precision ground truth data.

4.1 Experimental Setup and Data Acquisition

The experimental environment and camera configurations are summarized in Table 3.

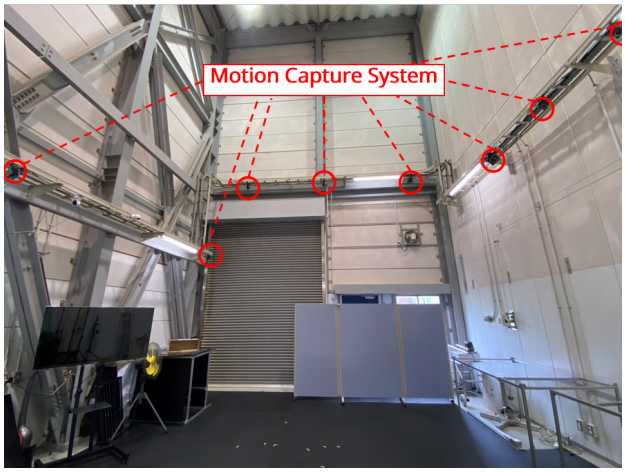
Table 3 – Experimental parameters and algorithm specifications

Parameter	Symbol	Unit	Value
Vehicle altitude	z	[m]	0.65
Slider target speed	u_{ref}	[m/s]	0.5
Attitude angles	(ϕ, θ)	[deg, deg]	$(0.0, 0.0)$
Cam 1 mounting config.	$(r_{c1}, \gamma_{c1}, \beta_{c1})$	[m, deg, deg]	$(0.06, -45, 90)$
Cam 1 optical axis	$(\gamma_{op,1}, \beta_{op,1})$	[deg, deg]	$(-45, 130)$
Cam 2 mounting config.	$(r_{c2}, \gamma_{c2}, \beta_{c2})$	[m, deg, deg]	$(0.06, -90, 90)$
Cam 2 optical axis	$(\gamma_{op,2}, \beta_{op,2})$	[deg, deg]	$(-90, 180)$
Field of View (H $\times$ V)	FOV	[deg]	26.5 $\times$ 20.0
Image resolution (H $\times$ V)	-	[pixels]	640 $\times$ 480
Sampling rate	f_s	[Hz]	60

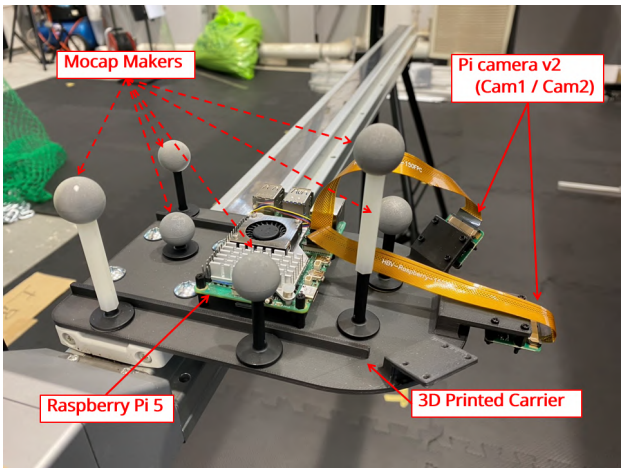
The sensor payload, equipped with a Raspberry Pi 5 and two Raspberry Pi Camera Module v2 units, is transported along the slider at a constant reference speed of 0.5 m/s ($u_{ref} = 0.5$ m/s). The vehicle altitude is maintained at $z = 0.65$ m keeping a level attitude ($\phi = 0, \theta = 0$). Two cameras are mounted

at positions located 0.06 m from the center of the payload, with azimuth/elevation angles of $(-45°, 90°)$ and $(-90°, 90°)$, respectively. Their optical axes are oriented at $(-45°, 130°)$ and $(-90°, 180°)$, respectively. Each camera has a field of view of $26.5°$ in the horizontal direction and $20.0°$ in the vertical direction, and assumes an intentionally downscaled image resolution of 640×480 pixels to enable real-time processing on the onboard edge computing device. Based on this geometry, dense optic flow is extracted using the DIS algorithm [13] and downsampled to a grid of 30×20 points in its flat image plane. To prevent motion blur and robustly track continuous visual features during the low-altitude transit, the sampling rate of the optic flow is strictly maintained at 60 Hz, while ground truth data is recorded using an OptiTrack motion capture system at 120 Hz.

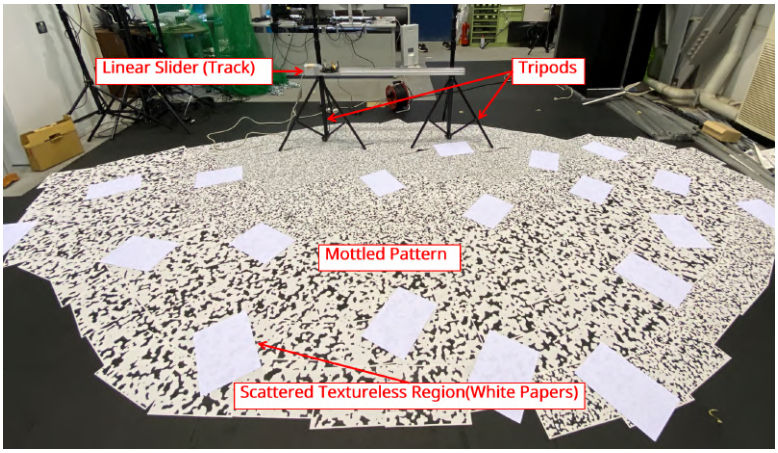
Figure 5 provides an overview of the experimental setup. The floor is covered with a specialized black-and-white mottled pattern whose size increases proportionally with the distance from the cameras to maintain a consistent signal-to-noise ratio. Moreover, multiple textureless white papers are scattered across the optical field to intentionally introduce severe and intermittent observation errors, mimicking featureless regions. The proposed framework is applied to the optic flow obtained under these challenging conditions.



(a) Laboratory environment



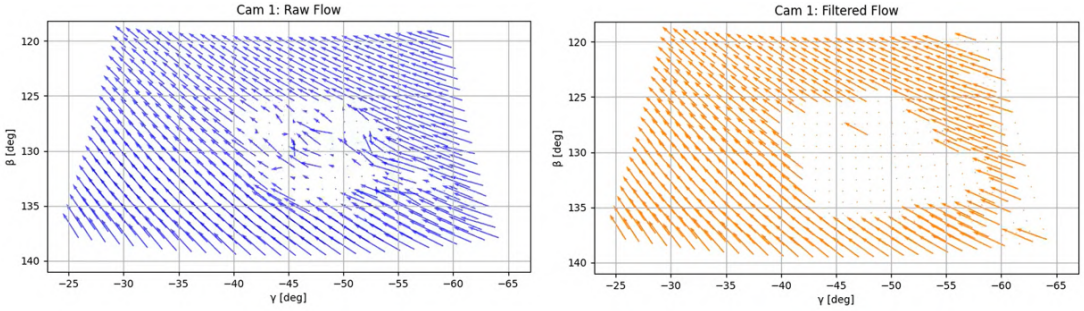
(b) Sensor payload on the carrier



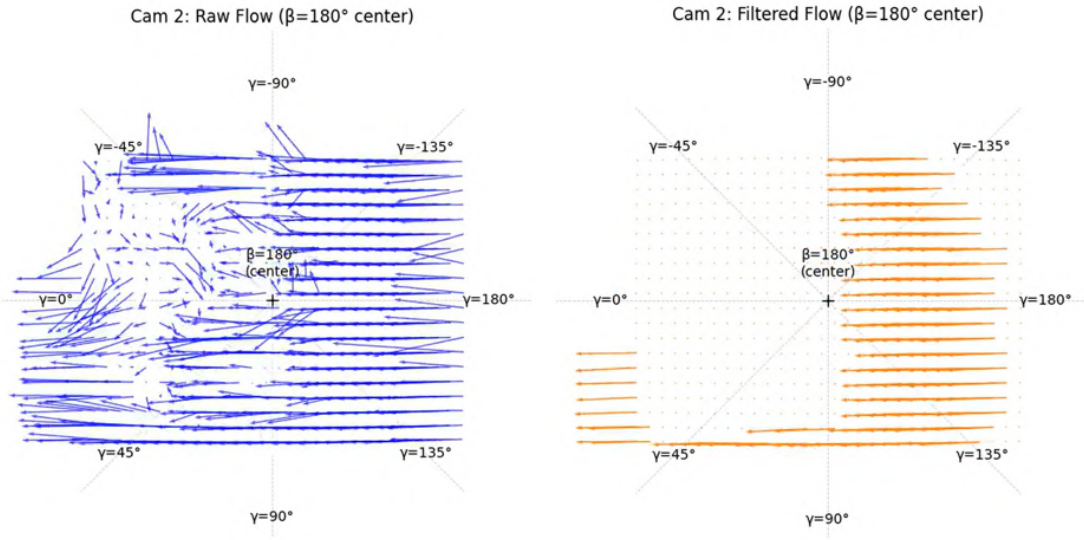
(c) Mottled floor texture with scattered textureless regions

Figure 5 – Experimental setup: (a) Overview of the laboratory, (b) the custom sensor payload comprising a Raspberry Pi 5 and dual camera modules, and (c) the mottled floor pattern interspersed with multiple white papers to simulate scattered visual disturbances.

4.2 Visual Analysis of Outlier Rejection Performance



(a) Cam 1 (Forward-downward view): Raw flow (left) vs. Filtered flow (right)



(b) Cam 2 (Nadir-downward view): Raw flow (left) vs. Filtered flow (right)

Figure 6 – Visualization of the spatial masking mechanism during the transit over a textureless region. In the raw flow fields (blue), the lack of texture causes the DIS algorithm to output erratic vectors. By applying the proposed method, these unreliable observations are effectively rejected, retaining only the valid flow vectors (orange) that are consistent with the actual vehicle motion.

Figure 6 shows the optic flow fields mapped onto the spherical coordinate system for Cam 1 (forward-downward) and Cam 2 (nadir-downward) during the visual disturbance. In the raw flow plots (blue vectors), the central regions corresponding to the white paper exhibit erratic or near-zero flow vectors. Because the DIS algorithm [13] relies on spatial intensity gradients, it fails to generate meaningful motion cues on these textureless surfaces. Consequently, these corrupted vectors completely contradict the actual 0.5 m/s translational motion of the slider and significantly distort the ideal flow field.
In the filtered flow plots (orange vectors), the proposed spatial mask effectively isolates and removes these unreliable vectors. As seen particularly in the Cam 2 results (Figure 6b), a substantial portion of the field of view is aggressively discarded based on the motion prediction. Despite this reduction in the number of spatial constraints, the remaining valid vectors consistently represent the correct motion state. This rigorous rejection guarantees that the unified state estimation is mathematically shielded from corrupted observations, establishing a reliable foundation for accurate motion estimation across all variables.
The fundamental origin of these observation errors is directly visible in the image plane. Figure 7 displays the raw optic flow overlaid on the actual camera frames. The lack of texture on the white paper prevents the extraction of any reliable motion vectors, highlighting the critical necessity of actively rejecting these regions before WFI of optic flow phase. By filtering out these anomalies, the proposed framework maintains strict navigation integrity.

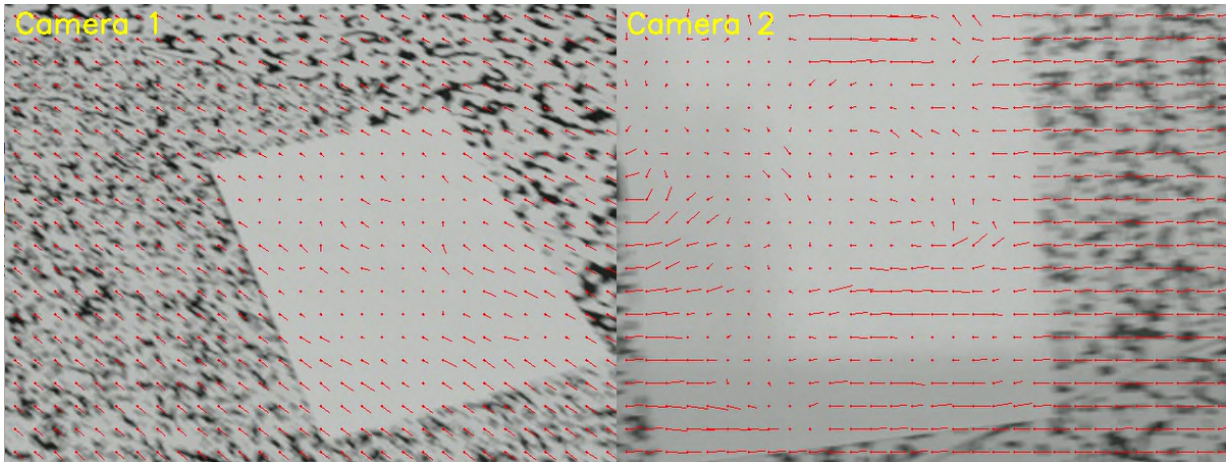


Figure 7 – Optic flow vectors overlaid on actual camera frames during the transit over a scattered white paper. The lack of texture prevents the extraction of valid motion cues, highlighting the necessity of active outlier rejection.

4.3 Evaluation of Motion Estimation Accuracy

Figure 8 presents the time-series estimation results during a single transit over the scattered white papers. For clarity, the specific data segment encompassing this transit is extracted from the continuous recording, and the time axis is realigned to start from $t = 0$ s.

As observed in the figure, the Original method suffers from severe oscillations and massive deviations from the true values across multiple axes during the transit over the textureless regions. For instance, the vertical velocity (w) exhibits abrupt transient errors exceeding 0.7 m/s, and the angular velocities show significant instability. These extreme errors, caused by the spatial integration of kinematically inconsistent flow vectors, would be fatal for maintaining the stable flight control of an actual MAV.

In contrast, the Proposed method demonstrates highly robust tracking performance across all six motion variables. By actively masking the corrupted flow vectors, as visually confirmed in Section 4.2, the proposed framework effectively suppresses the large transient deviations. It maintains a stable profile that closely follows the ground truth throughout the disturbance for both the translational and rotational components.

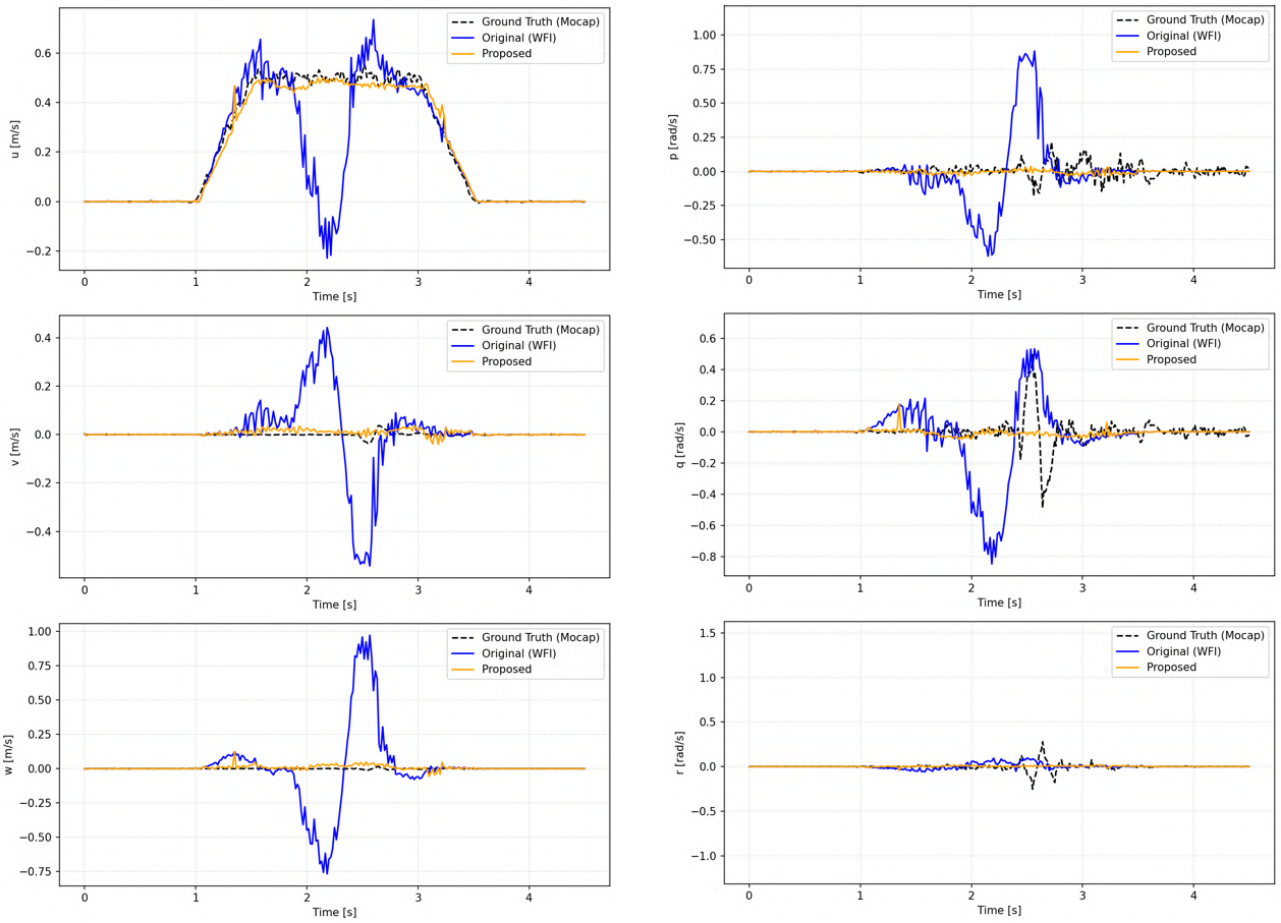
To provide a quantitative assessment, the Root Mean Square Error (RMSE) was calculated for the entire transit duration encompassing the visual disturbances. The results and the corresponding error reduction rates (defined in Eq.(11)) are summarized in Table 4.

Table 4 – Quantitative comparison of experimental estimation errors (RMSE)

Variable	Unit	Original RMSE	Proposed RMSE	Reduction Rate [%]
u	[m/s]	0.1652	0.0280	83.1
v	[m/s]	0.1416	0.0142	90.0
w	[m/s]	0.2464	0.0168	93.2
p	[rad/s]	0.2279	0.0452	80.2
q	[rad/s]	0.2125	0.0834	60.8
r	[rad/s]	0.0477	0.0355	25.6

As shown in Table 4, the proposed method achieved a substantial reduction in RMSE across all motion variables. Notably, the translational velocities (u, v, w) exhibited improvement rates exceeding 83%, effectively maintaining the absolute velocity error below 0.03 m/s even under severe visual degradation.

Furthermore, the estimation accuracy of the rotational components (p, q, r) was also consistently improved. While the improvement rate of r was relatively modest at 25.6%, this is attributable to the fact that the standard method already yields small absolute errors in r estimation, leaving limited margin for further reduction. In contrast, p and q achieved over 60% error reduction. Consistent with the



(a) Comparison of translational velocities (u, v, w)

(b) Comparison of angular velocities (p, q, r)

Figure 8 – Time-series comparison of the estimated motion variables during the transit over scattered featureless regions. For a kinematically equivalent visual comparison, the vertical axis limits of the angular velocity plots are scaled by $1/z$ (where altitude $z = 0.65$ m) relative to the translational velocity plots. The Proposed method (orange solid line) effectively suppresses the large transient deviations seen in the standard WFI of optic flow algorithm (blue solid line) and stably tracks the ground truth (black dashed line) across all motion variables.

numerical simulation results presented in Section 3, this comprehensive improvement quantitatively confirms that excluding unreliable flow vectors fundamentally enhances the numerical stability of the entire WFI of optic flow algorism. These results validate the practical effectiveness of the proposed framework for reliable autonomous MAV navigation in real-world cluttered environments.

5. Conclusion

This study proposed an outlier rejection framework in the WFI of optic flow, based on excluding unreliable flow vectors to enhance the robustness of motion estimation under severe practical visual disturbances. Through numerical simulations and physical slider experiments, the proposed method successfully suppressed large transient deviations caused by featureless regions. By actively isolating and removing these kinematically inconsistent observations, the framework achieved highly accurate estimation across all motion variables, substantially reducing both translational and rotational errors. This comprehensive robustness enhancement establishes a critical foundation for reliable state estimation in degraded visual environments. Future work will focus on validating the estimation performance using visual data captured by an actual MAV during flights with visual disturbances. Ultimately, we intend to apply this framework to a real-time control system for MAVs, with the goal of realizing robust autonomous flight in cluttered and textureless indoor environments.

6. Contact Author Email Address

mailto: yanagisawa.masaya.769@s.kyushu-u.ac.jp

7. Copyright Statement

The authors confirm that they, and/or their company or organization, hold copyright on all of the original material included in this paper. The authors also confirm that they have obtained permission, from the copyright holder of any third party material included in this paper, to publish it as part of their paper. The authors confirm that they give permission, or have obtained permission from the copyright holder of this paper, for the publication and distribution of this paper as part of the ICAS proceedings or as individual off-prints from the proceedings.

References

[1] Yaoqiang Pan, Kewei Hu, Xiao Huang, Wei Ying, Xiaoxuan Xie, Yue Ma, Naizhong Zhang, and Hanwen Kang. Developing smart mavs for autonomous inspection in gps-denied constructions. *arXiv preprint arXiv:2408.06030*, 2024.

[2] Yunfan Ren, Yixi Cai, Haotian Li, Nan Chen, Fangcheng Zhu, Longji Yin, Fanze Kong, Rundong Li, and Fu Zhang. A survey on LiDAR-based autonomous aerial vehicles. *IEEE/ASME Transactions on Mechatronics*, 2025. Accepted for publication.

[3] K. N. McGuire, C. De Wagter, K. Tuyls, H. J. Kappen, and G. C. H. E. de Croon. Minimal navigation solution for a swarm of tiny flying robots to explore an unknown environment. *Science Robotics*, 4(35):eaaw9710, 2019.

[4] Carlos Campos, Richard Elvira, Juan J. Gómez Rodríguez, José M. M. Montiel, and Juan D. Tardós. ORB-SLAM3: An accurate open-source library for visual, visual–inertial, and multi-map SLAM. *IEEE Transactions on Robotics*, 37(6):1874–1890, 2021.

[5] Julien Bailly and Guido de Croon. Antflie: Frugal visual teach and repeat on narrow FoV micro-drones. *arXiv preprint arXiv:2509.10730*, 2025.

[6] M. A. Shoemaker and S. Hokamoto. Comparison of integrated and nonintegrated wide-field optic flow for vehicle navigation. *Journal of Guidance, Control, and Dynamics*, 36:710–720, 2013.

[7] J. R. Serres. Taking inspiration from flying insects to navigate inside buildings. In *Interdisciplinary Expansions in Engineering and Design With the Power of Biomimicry*, chapter 3. InTechOpen, 2018.

[8] J. S. Humbert. *Bio-inspired visuomotor convergence in navigation and flight control systems*. PhD thesis, California Institute of Technology, 2005.

[9] Naoto Kobayashi, Masataka Oishi, Yutaka Kinjo, and Shinji Hokamoto. Experimental verification of wide-field-integration of optic flow for state estimation. *Transactions of the Japan Society for Aeronautical and Space Sciences, Aerospace Technology Japan*, 14(ists30):Pd_63–Pd_68, 2016.

[10] Joseph Conroy, Gregory Gremillion, Badri Ranganathan, and J. Sean Humbert. Implementation of wide-field integration of optic flow for autonomous quadrotor navigation. *Autonomous Robots*, 27(3):189–198, 2009.

[11] J. Sean Humbert and Andrew M. Hyslop. Bioinspired visuomotor convergence. *IEEE Transactions on Robotics*, 26(1):121–130, 2010.

[12] Ryuichiro Okabayashi, Riko Shirahase, Mai Bando, and Shinji Hokamoto. Practical motion estimation for micro aerial vehicles using wide-field-integration of optic flow. In *AIAA SCITECH 2026 Forum*, 2026.

[13] Till Kroeger, Radu Timofte, Dengxin Dai, and Luc Van Gool. Fast optical flow using dense inverse search. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 471–488, 2016.