

<https://doi.org/10.31891/csit-2026-3-12>

Anatolii ZHARKOV

PhD student,
Vinnytsia National Technical University
Vinnytsia, Ukraine
<https://orcid.org/0009-0005-9857-644X>
e-mail: fkca.lakitjav@gmail.com

Roman MASLII

PhD, Associate Professor,
Vinnytsia National Technical University
Vinnytsia, Ukraine
<https://orcid.org/0000-0003-3021-4328>
e-mail: maslij.r.v@vntu.edu.ua

Received: 04/08/2026

Accepted: 11/09/2026

Published: 30/09/2026

© Copyright
2026 by the author(s)



This is an Open Access article distributed
under the terms of the [Creative Commons
CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)

UDC 004.896:681.51

RIGHT-INVARIANT STEREO VISUAL- INERTIAL ODOMETRY: REJECTION THRESHOLDS AND NOISE SCALE

A mathematical model is presented of a visual-inertial odometry filter built on a right-invariant error on the matrix Lie group $SE_2(3)$ with appended inertial measurement unit biases, with landmarks eliminated by projecting the residual onto the left null space of the Jacobian and with measurements rejected by a normalised Mahalanobis distance. For every element of the model the transition matrices, Jacobians and retraction are given in both error conventions, the right-invariant and the classical additive one, which makes it possible to isolate the effect of the convention itself under otherwise identical conditions.

On five EuRoC MH sequences with four front ends arranged in a 2×2 design (two ways of tracking a point $\times$ two detectors: the learned XFeat and the classical ORB), the separate and joint contributions of two factors are measured: the error convention and the form of the rejection rule. The widely used constant threshold is shown to correspond to a χ^2 quantile for a two-component measurement and not to agree with the actual dimension of the projected residual. Replacing it by a quantile threshold reduces the mean position error by 21.1, 23.4 and 20.2 % on three front ends out of four: the best result of the grid is 0.136 m against 0.177 m with the constant threshold and 0.222 m with no rejection at all. The contribution of the error convention is smaller, 2.4–10.0 %, and on ORB with descriptor matching the ratio of the two factors reverses. At the same time the same replacement, with the noise scale left unchanged, causes a loss of solution in one configuration out of twenty (644.9 m), which is removed by matching σ (0.436 m) rather than by relaxing the rule.

The scientific novelty consists in establishing that the form of the rejection threshold and the adopted measurement-noise scale are subject to joint rather than separate calibration: a threshold defined by a χ^2 quantile at the actual residual dimension needs no tuned constant, yet transferred without matching σ it costs the solution. The practical value lies in assigning the threshold without manual tuning and in the quantitatively stated limits of transferability of settings between front ends.

Keywords: visual-inertial odometry, invariant Kalman filter, Lie group, unmanned aerial vehicle, outlier rejection, EuRoC.

Introduction

Simultaneous localization and mapping remains one of the central problems of mobile robotics, and survey work records the shift of the field towards requirements of reliability and reproducibility of results [1]. Visual-inertial odometry (VIO) is the basic means of autonomous navigation for unmanned aerial vehicles where satellite navigation is unavailable: position, orientation and velocity are estimated from the readings of an inertial measurement unit (IMU) and a sequence of images, without an external positional reference [2]. The inertial channel provides a solution whose error grows without bound, and the visual channel is meant to bound that growth.

The dominant filter-based approach is the multi-state constraint Kalman filter (MSCKF) [3], in which landmarks do not enter the state but are eliminated at the update step, which gives a cost linear in the number of features. Independently of this, the theory of invariant filtering [4] has become established: the error is defined not additively but through the action of a Lie group, owing to which the linearized error dynamics ceases to depend on the estimate itself.

Both solutions are described in the literature. Far less studied is the question of how stable the gain from invariance is with respect to the rest of the pipeline, primarily with respect to the way features are obtained (the front end) and to the measurement rejection rule. If the advantage of a convention

changes sign when the front end is replaced, then a comparison carried out on a single configuration is not a ground for an architectural decision.

Analysis of recent research and publications

The multi-state constraint filter was proposed in [3]: a sliding window of cloned poses, a landmark eliminated by projecting the residual onto the left null space of the landmark Jacobian, and measurements rejected by a Mahalanobis test. This scheme underlies most modern filter-based implementations, in particular OpenVINS [5] and the stereo variant [6].

The theory of the invariant Kalman filter [4] shows that for systems on Lie groups whose dynamics are group-affine it is appropriate to define the error as a group element: its linearized dynamics then does not depend on how accurate the estimate currently is. For inertial navigation the corresponding group is $SE_2(3)$, which carries orientation, velocity and position at once. The IMU biases do not fit into the group structure, so they are appended to the state separately; such a construction is known as an “imperfect” invariant filter, and a residual dependence on the estimate is retained in the blocks associated with the biases.

Pre-integration of inertial measurements [7] makes it possible to reduce all IMU readings between two frames to a single relative increment together with its covariance, which makes the propagation step independent of the IMU rate.

The line is still active and not settled. An equivariant filter has been combined with the multi-state constraint structure [8], giving an alternative to the invariant formulation within the same elimination scheme, and consistency conditions for nonlinear estimators, including visual-inertial odometry, have been derived in a transformation-based framework [9]. It has also been argued that, once a reset step is included, the left- and right-invariant filters are in fact identical [10], which shows that the choice of convention is still under discussion rather than closed. This is an additional reason to measure its contribution empirically rather than to assume it.

The combination of these two solutions is already known: in [11] right-invariant error states are implemented on top of OpenVINS, with the features decoupled from covariance propagation, which preserves the computational advantage of the scheme with landmark elimination. This work therefore does not propose a new combination but measures the limits of applicability of a known one.

What the listed works have in common is that each fixes a single feature-tracking scheme, so they provide no data on the stability of their conclusions when the front end is replaced.

This work continues the authors’ own line. In [12] the suitability of the main Visual SLAM approaches for autonomous robot navigation was analyzed, and in [13] visual-inertial SLAM based on the extended Kalman filter was studied together with the effect of the ratio between motion-model and measurement noises on its accuracy. Here that line is developed in two directions: the classical error convention is replaced by the right-invariant one, and the effect of the settings is measured not on a single configuration but on a grid of four front ends.

Objectives and tasks

The objective of the work is to present the mathematical model of a right-invariant filter for stereo visual-inertial odometry in a completeness sufficient for reproduction, and to measure the separate and joint effect of two factors that are switched independently of each other within this model: the convention defining the state error and the rule for rejecting measurements by a normalized Mahalanobis distance.

To reach the objective the following tasks were solved:

1. to build a state model on the matrix Lie group $SE_2(3)$ with appended inertial measurement unit biases and to derive the transition matrices, measurement Jacobians and retraction in both error conventions, the right-invariant and the classical additive one;
2. to examine the agreement of the widely used constant rejection threshold with the dimension of the measurement and to build an alternative that agrees with it;
3. to measure the separate contribution of the error convention and of the form of the rejection rule under otherwise identical conditions on four front ends arranged in a 2×2 design;
4. to find out whether the conclusions obtained transfer between front-end configurations and between forms of the rejection rule itself.

Problem statement

Given a sequence of IMU readings $\{(\omega_k, a_k)\}$ and synchronised stereo images, the trajectory $\{(R_t, v_t, p_t)\}$ in a fixed frame has to be estimated together with the gyroscope and accelerometer biases b_g, b_a .

Within this work the model is built so that the factors are switches that change nothing else:

- the *error convention* $\varepsilon \in \{\text{right}, \text{additive}\}$ affects exactly three places: the transition matrix, the orientation block of the measurement Jacobian and the retraction;
- the *rejection rule* has three levels: off, the constant threshold $5.99m$ and the quantile threshold $\chi^2_{m,0.95}$; it affects exactly one place, the acceptance condition for the projected residual. Both non-zero forms use the same Mahalanobis distance and differ only in how the threshold is assigned.

The rest of the pipeline (pre-integration, triangulation, landmark elimination, window, noise covariances) stays bit-for-bit the same. This is the “all else equal” condition, without which measuring the difference between the conventions is meaningless.

Materials and methods

The state as an element of a matrix Lie group

The navigation state is written as a single matrix

$$X = \begin{bmatrix} R & v & p \\ 0_{1 \times 3} & 1 & 0 \\ 0_{1 \times 3} & 0 & 1 \end{bmatrix} \in SE_2(3), R \in SO(3), v, p \in R^3, \quad (1)$$

where matrix multiplication reproduces the composition of transformations. The full filter state

$$S = (X, b_g, b_a, \{(R_i, p_i)\}_{i=1}^N) \quad (2)$$

contains, besides X and the biases, N cloned poses of a sliding window. Landmarks are not in the state, because they are eliminated at the update step by projecting the residual onto the null space, as described below; the state dimension $15 + 6N$ therefore does not depend on the number of observed points.

The Lie algebra $se_2(3)$ is identified with R^9 through $\xi = (\xi_\theta, \xi_v, \xi_p)$, and the exponential has the closed form

$$\exp(\xi) = \begin{bmatrix} \exp_{SO(3)}(\xi_\theta) & J(\xi_\theta)\xi_v & J(\xi_\theta)\xi_p \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad (3)$$

where $\exp_{SO(3)}$ is the Rodrigues formula and J is its left Jacobian.

Pre-integration of inertial measurements

Between adjacent frames k and $k + 1$ the IMU readings reduce to three increments [7]

$$\Delta R_k = \prod_{t_k \leq \tau < t_{k+1}} \exp_{SO(3)}((\omega_\tau - b_g)\delta t), \Delta v_k = \sum_\tau \Delta R_{k,\tau} (a_\tau - b_a) \delta t, \quad (4)$$

$$\Delta p_k = \sum_\tau [\Delta v_{k,\tau} \delta t + 1/2 \Delta R_{k,\tau} (a_\tau - b_a) \delta t^2], \quad (5)$$

which do not depend on the state at t_k and are therefore computed once. The state is propagated as

$$R_{k+1} = R_k \Delta R_k, v_{k+1} = v_k + g \Delta t_k + R_k \Delta v_k, p_{k+1} = p_k + v_k \Delta t_k + 1/2 g \Delta t_k^2 + R_k \Delta p_k. \quad (6)$$

The dependence of the increments on the biases is accounted for to first order through the Jacobians $\partial \Delta / \partial b$, so recomputation when the biases are updated does not require a second pass over the readings. The covariance of the segment is transported into the fixed frame by the adjoint map Ad_X .

The right-invariant error and its dynamics

The error is defined *right-invariantly*, that is, as a group element

$$\eta = X \hat{X}^{-1} = \exp(\xi), \xi = (\delta\theta, \delta v, \delta p) \in R^9, \quad (7)$$

in contrast to the additive definition, in which the orientation, velocity and position errors are treated as independent vectors. Differentiating (7) along the dynamics $\dot{v} = Ra + g$ gives the linearised system

$$\delta\dot{\theta} = 0, \delta\dot{v} = [g]_\times \delta\theta, \delta\dot{p} = \delta v, \quad (8)$$

which contains *none of the estimates* $\hat{R}$, $\hat{v}$, $\hat{p}$ and none of the readings $\hat{a}$, $\hat{\omega}$. In matrix form the navigation block is

$$A_{\text{nav}} = \begin{bmatrix} 0 & 0 & 0 \\ [g]_\times & 0 & 0 \\ 0 & I & 0 \end{bmatrix}, A_{\text{nav}}^3 = 0, \quad (9)$$

so the transition matrix is computed exactly as a second-degree polynomial, $\Phi = I + A\Delta t + 1/2 A^2 \Delta t^2$, with no truncation of the series.

The columns that couple the error to the bias increments do depend on the estimate:

$$A_{\delta\theta, b_g} = -\hat{R}, A_{\delta v, b_g} = -[\hat{v}]_\times \hat{R}, A_{\delta p, b_g} = -[\hat{p}]_\times \hat{R}, A_{\delta v, b_a} = -\hat{R}. \quad (10)$$

This is exactly the residual dependence on the state that the “imperfect” construction of the invariant filter cannot remove: a bias is not an element of $SE_2(3)$, and appending it to the state inevitably brings $\hat{R}$ into the dynamics. To claim complete independence of the transition matrix from the estimate, as is sometimes done, would be inaccurate; what is independent is the navigation block (8), and it is this block that dominates.

The window clones are frozen at the propagation step, so the full transition matrix is block-diagonal, $\Psi = \text{diag}(\Phi_{15 \times 15}, I_{6N})$, and the multiplication is carried out blockwise.

The classical additive convention

For comparison under otherwise identical conditions a second convention is implemented, in which the error is defined additively and the rotation is applied by right multiplication in the body frame: $R = \hat{R} \exp_{SO(3)}(\delta\theta)$, $v = \hat{v} + \delta v$, $p = \hat{p} + \delta p$. Its linearised dynamics

$$\delta\dot{\theta} = -[\hat{\omega}]_\times \delta\theta - \delta b_g, \delta\dot{v} = -\hat{R}[\hat{a}]_\times \delta\theta - \hat{R}\delta b_a, \delta\dot{p} = \delta v \quad (11)$$

contains both the orientation estimate $\hat{R}$ and the readings $\hat{a}$, $\hat{\omega}$ themselves. Comparing (8) and (11) shows the essence of the difference in a single line: in the first case the matrix is constant, in the second it is rebuilt at every step from the current estimate. The consequence is that when the orientation error is large the additive convention linearises the dynamics about a wrong point, which the right convention does not do.

Since A in (11) is not nilpotent, the transition matrix is computed by a truncated series $\Phi_{\text{ekf}} = I + A\Delta t + 1/2 (A\Delta t)^2 + 1/6 (A\Delta t)^3$.

The stereo camera measurement model

A landmark l in the fixed frame, observed from a cloned pose (R_j, p_j) , is transformed into the camera frame by a rigid transformation (R_{ci}, t_{ci}) :

$$p_j^c = R_{ci} \left(R_j^T (l - p_j) \right) + t_{ci} = (X, Y, Z)^T, \quad (12)$$

and the measurement is four-dimensional, two coordinates in the left camera and two in the right:

$$z_j = (f_x X/Z + c_x, f_y Y/Z + c_y, f_x X - b/Z + c_x, f_y Y/Z + c_y)^T, \quad (13)$$

where b is the stereo baseline. Let $J_\pi = \partial z / \partial p^c$ and $B_j = J_\pi R_{ci} R_j^T$. The landmark Jacobian is the same for both conventions, $H_f^{(j)} = B_j$, whereas the state Jacobian differs:

$$H_x^{(j)} = [B_j[l]_x - B_j](\text{right}), H_x^{(j)} = [J_\pi R_{ci} [\hat{h}_j]_x - B_j](\text{additive}), \quad (14)$$

where $\hat{h}_j = R_j^T (l - p_j)$. The difference is substantive: in the right convention the orientation column is proportional to $[l]_x$, that is, to the *position of the landmark in the fixed frame*, whereas in the additive one it is $[\hat{h}_j]_x$, the vector *from the camera to the landmark*. In the second case the sensitivity is bounded by the observation range, in the first it is not.

Landmark elimination by projection onto the null space

Stacking the rows of (14) over all M views of a feature gives

$$r = H_x \mathbf{x} + H_f \mathbf{t} + n, n \sim N(0, \sigma^2 I). \quad (15)$$

The landmark $\mathbf{t}$ is unknown and is not contained in the state. Let A be an orthonormal basis of the left null space of H_f , that is, $A^T H_f = 0$; it is obtained from the singular value decomposition $H_f = U \Sigma V^T$ as the columns of U corresponding to zero singular values. Multiplying from the left by A^T gives

$$r_n = A^T r = H_n \mathbf{x} + A^T n, H_n = A^T H_x, \quad (16)$$

a system from which the landmark has disappeared exactly rather than approximately. Since A is orthonormal, the noise covariance keeps the form $\sigma^2 I$, so the projection does not require recomputing the weighting matrix.

The geometric meaning of (16) matters for the next subsection: $A^T r$ is exactly the part of the residual that *cannot be explained by any landmark position*. The reprojection residual does not have this property: a feature with a wrong correspondence or on a moving object is triangulated successfully, only not at the right point in space, after which its reprojections are consistent across views.

The number of rows of r_n equals $NR \cdot M - 3$ and for long tracks exceeds the state dimension; QR compression is then applied, $H_n = [Q_1 Q_2] \begin{bmatrix} T \\ 0 \end{bmatrix}$, and only T and $Q_1^T r_n$ enter the update, which does not change the result but makes the update cost independent of the track length.

Rejection by a normalized Mahalanobis distance

The projected residual (16) has its own covariance

$$S = H_n P H_n^T + \sigma^2 I, \quad (17)$$

which accounts both for the state uncertainty and for the measurement noise. A feature is accepted into the update if

$$q = \frac{r_n^T S^{-1} r_n}{\dim r_n} \leq \gamma, \gamma = \chi_{2,0.95}^2 = 5.99, \quad (18)$$

and is rejected otherwise. By its name this is a *normalized Mahalanobis-distance gate* with a fixed threshold of 5.99, not a classical χ^2 test.

Two forms of the same statistic are used below, and it is worth separating them once. The *unnormalized* Mahalanobis statistic is $d = r_n^T S^{-1} r_n$, which under a correct model follows χ_m^2 ; the *normalized* one is $q = d/m$, used in (18) so that a single number can serve tracks of different length. The rule $q \leq 5.99$ is therefore equivalent to $d \leq 5.99m$, and it is this equivalent unnormalised threshold that is comparable with the classical $\chi_{m,0.95}^2$. All comparisons of thresholds below are stated for d .

Agreement of the threshold with the measurement dimension. The number 5.99 is the quantile $\chi_{2,0.95}^2$, that is, a value for a measurement with *two* components, a monocular one. In the stereo setting there are four rows per view, and eliminating the landmark removes three degrees of freedom, so

$$m = \dim r_n = 4M - 3, \quad (19)$$

where M is the number of views in the track. The value $m = 2$ is unreachable by construction: already for the shortest usable track, $M = 2$, we have $m = 5$. The classical criterion of the same confidence level takes the threshold *at the actual dimension*, $r_n^T S^{-1} r_n \leq \chi_{m,0.95}^2$, and since the quantile grows with m more slowly than a linear function, the constant threshold $5.99m$ turns out to be substantially softer (Table 1).

Table 1.

The constant threshold against the quantile one at the actual residual dimension

M	m	$\chi^2_{m,0.95}$	$5.99m$	softer by	actual level
2	5	11.07	29.95	2.7	$1.5 \cdot 10^{-5}$
4	13	22.36	77.87	3.5	$2.8 \cdot 10^{-11}$
8	29	42.56	173.71	4.1	$1.4 \cdot 10^{-22}$
12	45	61.66	269.55	4.4	$9.0 \cdot 10^{-34}$

The last column gives the probability of rejecting a correct measurement, that is, the significance level to which the constant threshold actually corresponds. Since the number 5.99 is taken from a quantile of level 0.95, the criterion ought to reject about 5 % of correct measurements; in fact it rejects from $1.5 \cdot 10^{-5}$ at $M = 2$ down to $9.0 \cdot 10^{-34}$ at $M = 12$, that is, the longer the track, the less it rejects. The constant in the numerator therefore makes the criterion not only softer but also *non-uniform*: the strictness of the test depends on how many frames the feature has lived. The quantile form removes this arbitrariness: the threshold needs no constant at all and agrees with the track length by itself, through (19).

Below the two forms are compared as two levels of one factor: the constant threshold $5.99m$ as common practice, and the quantile threshold $\chi^2_{m,0.95}$ as a criterion that agrees with the dimension.

The criterion (18) acts *after* landmark elimination, that is, on the same quantity $A^T r$ rather than on the reprojection residual.

Update and retraction

The accepted rows are stacked into a common system and the standard Kalman update is performed

$$K = PH^T(HPH^T + \sigma^2 I)^{-1}, \delta x = Kr, P \leftarrow (I - KH)P(I - KH)^T + K\sigma^2 K^T, \quad (20)$$

where the Joseph form is used to preserve symmetry and positive definiteness.

The way the increment δx is applied is determined by the convention. In the right convention it is a multiplication *from the left*, consistent with the definition (7):

$$\hat{X} \leftarrow \exp(\delta x_{1:9})\hat{X}, \text{ that is } \hat{R} \leftarrow \delta R \hat{R}, \hat{v} \leftarrow \delta R \hat{v} + \delta v, \hat{p} \leftarrow \delta R \hat{p} + \delta p, \quad (21)$$

and the window clones are retracted in the same way, because their error is defined in the same frame. In the additive convention the increment is applied component wise, $\hat{R} \leftarrow \hat{R} \exp_{SO(3)}(\delta \theta)$, $\hat{v} \leftarrow \hat{v} + \delta v$, $\hat{p} \leftarrow \hat{p} + \delta p$. The biases are updated additively in both cases.

Mixing the conventions in even one of the three places (transition matrix, Jacobian, retraction) leads to a silent divergence between the estimate and the covariance that is not revealed by any quantity other than consistency measures. In the implementation, therefore, the two branches are switched by a single flag and partial combinations are forbidden.

Computational cost

Since landmarks do not enter the state, the dimension of P equals $15 + 6N$ and does not depend on the number of features, while QR compression removes the dependence on the track length. Both conventions cost the same: the difference between them is in the content of the matrices, not in their sizes. The gate requires solving a system of size $\dim r_n$ per feature, but it also reduces the number of update rows.

Experimental procedure

Data. Five EuRoC MH sequences (MH01–MH05) [14] are used: flights of a micro quadrotor in a machine hall with a reference trajectory from a laser tracker and a motion capture system. The input consists of 20 Hz stereo images and 200 Hz IMU readings; the filter receives no additional information about position, and the initialization is self-contained.

Front ends. Four ways of obtaining features are compared, arranged in a 2×2 design: two ways of tracking a point between frames, each with two detectors.

- tracking by optical flow (Lucas-Kanade): XFeat-1024+KLT with the learned detector [15] and ORB-1024+KLT with the classical one [16];
- tracking by descriptor matching: XFeat-1024+desc and ORB-1024+desc.

The design is needed to test transferability: by varying only the detector or only the tracking method one cannot tell which of the two factors determines the behavior. Here *both* detectors are present in *both* branches, so neither factor is tied to the other. Shi-Tomasi corners [17], often taken as the classical front end, do not fit into this design structurally: they have no descriptor, so they cannot be crossed with the second tracking method; a control run with them is mentioned in the results.

Metric. Let $\{p_k\}_{k=1}^K$ be the reference positions and $\{\hat{p}_k\}$ the estimated ones. Since the estimator has no external reference, the trajectories are first aligned by a rigid transformation that minimises the sum of squared deviations:

$$(R_a, t_a) = \arg \min_{R \in SO(3), t \in \mathbb{R}^3} \sum_{k=1}^K \|p_k - (R\hat{p}_k + t)\|^2, \quad (22)$$

after which the metric is the root-mean-square norm of the residual deviations

$$ATE3D = (1/K \sum_{k=1}^K \|p_k - (R_a \hat{p}_k + t_a)\|^2)^{1/2}. \quad (23)$$

The scale in (22) is not fitted: the stereo baseline sets the metric scale, so fitting it would remove part of the error that the metric is meant to measure.

Configuration. A window of $N = 30$ clones and $\sigma = 1.5$ px for all regimes without exception. A common σ is needed because tuning σ separately for each cell would change two factors at once, and the separate contribution of the rejection rule would be unmeasurable. The estimator is deterministic, so no stochastic repetition of the runs was applied. The only departure from the common σ is deliberate and described separately: in the configuration where the quantile threshold loses the solution, σ is varied in order to find out whether the cause is the form of the rule or its disagreement with the noise scale.

Results and discussion

The results of all one hundred and twenty cells are given in Table 2 and in Fig. 1. The rejection rule is presented at three levels: no gate, the constant threshold $5.99m$ and the quantile threshold $\chi^2_{m,0.95}$, both non-zero forms acting on the normalised Mahalanobis distance and differing only in how the threshold is assigned. Every number is a single run of a complete sequence at $\sigma = 1.5$ px and $N = 30$; the best value in a row is set in bold. The mean for ORB + KLT is marked with an asterisk and computed over four sequences: the cell in which the filter loses the solution is excluded from the averaging, because otherwise the mean would lose its meaning. In Fig. 1 the two left panels correspond to tracking by optical flow and the two right ones to descriptor matching; the scale is logarithmic and clipped at 3 m, with values above it labelled by number.

Table 2.

ATE3D (m) on EuRoC MH by error convention and by the form of the rejection rule

Front end	Seq.	right-invariant			additive (EKF)		
		none	$5.99m$	χ^2_m	none	$5.99m$	χ^2_m
XFeat-1024 + KLT	MH01	0.492	0.415	0.315	0.527	0.422	0.322
	MH02	0.186	0.191	0.148	0.194	0.203	0.162
	MH03	0.355	0.218	0.201	0.392	0.251	0.230
	MH04	0.545	0.333	0.278	0.574	0.337	0.265
	MH05	0.592	0.528	0.387	0.622	0.544	0.402
	<i>середнє</i>	0.434	0.337	0.266	0.462	0.351	0.276
ORB-1024 + KLT	MH01	0.550	0.487	645	0.569	0.496	0.383
	MH02	0.405	0.362	0.267	0.401	0.376	0.272
	MH03	0.247	0.214	0.175	0.294	0.277	0.206
	MH04	0.592	0.283	0.212	0.573	0.362	0.218
	MH05	0.577	0.564	0.481	0.494	0.536	0.498
	<i>середнє*</i>	0.474	0.382	0.284	0.466	0.409	0.315
XFeat-1024 + desc	MH01	0.120	0.114	0.126	0.110	0.105	0.128
	MH02	0.096	0.072	0.069	0.096	0.072	0.069
	MH03	0.129	0.106	0.134	0.156	0.124	0.159
	MH04	0.395	0.258	0.149	0.426	0.285	0.162
	MH05	0.369	0.335	0.201	0.349	0.320	0.213
	<i>середнє</i>	0.222	0.177	0.136	0.227	0.181	0.146
ORB-1024 + desc	MH01	0.148	0.106	0.101	0.165	0.124	0.119
	MH02	0.120	0.112	0.112	0.135	0.119	0.119
	MH03	0.214	0.216	0.224	0.232	0.232	0.237
	MH04	0.168	0.167	0.168	0.169	0.171	0.173
	MH05	0.300	0.301	0.289	0.354	0.356	0.344
	<i>середнє</i>	0.190	0.180	0.179	0.211	0.200	0.198

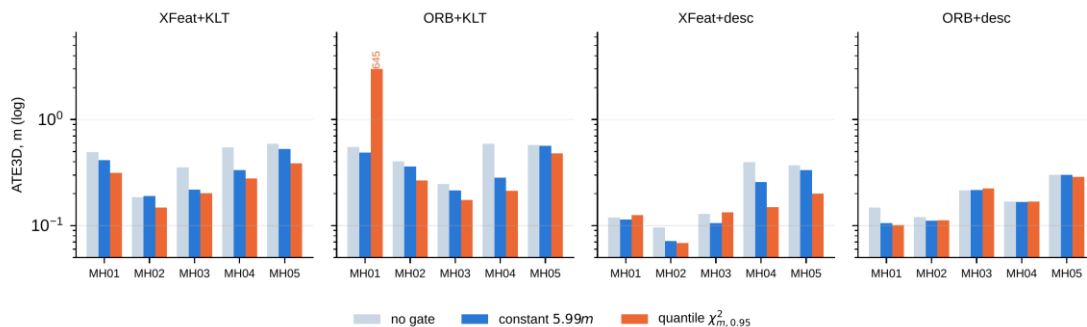


Fig. 1. ATE3D by sequence for three forms of the rejection rule, right-invariant convention

The form of the threshold matters more than its presence. Moving from the constant threshold to the quantile one gives more than introducing the gate did in the first place: the mean error falls by 21.1 % (XFeat + KLT),

23.4 % (XFeat+desc) and 20.2 % (ORB+KLT, without the diverging cell, see below), whereas on ORB+desc the change of form gives nothing (0.8 %). The latter agrees with the rest of the measurements: on this front end the gate rejects the least anyway, so a stricter threshold has nothing to remove. The constant 5.99, which corresponds to a two-component measurement, therefore costs about a fifth of the accuracy wherever rejection works at all.

The gain is uniform on the flow-tracked branch and concentrated on the other one. A mean over five sequences can hide the distribution, so Table 3 gives the median and the range together with the number of cells in which the quantile threshold wins. On the two KLT front ends the improvement is present in every cell and the median moves with the mean, by 16.6 and 25.7 %. On XFeat+desc the picture differs: the quantile threshold wins in three cells out of five, the mean gain of 23.4 % comes from the two hardest sequences (MH04 and MH05, -42 and -40 %), and the median rises by 17 %. On ORB+desc the two forms are indistinguishable cell by cell (-5 to +4 %). The mean improvement should therefore be read as uniform on the flow-tracked branch and as concentrated on the difficult sequences elsewhere.

Table 3.

Median ATE3D (m) over the five sequences, right-invariant convention			
front end	constant 5.99m	quantile $\chi^2_{m,0.95}$	cells won
XFeat-1024 + KLT	0.333 [0.191–0.528]	0.278 [0.148–0.387]	5/5
ORB-1024 + KLT*	0.322 [0.214–0.564]	0.239 [0.175–0.481]	4/4
XFeat-1024 + desc	0.114 [0.072–0.335]	0.134 [0.069–0.201]	3/5
ORB-1024 + desc	0.167 [0.106–0.301]	0.168 [0.101–0.289]	2/5

But the stricter threshold has a limit, and it is abrupt. In the cell ORB+KLT/MH01 with the right convention the quantile threshold gives 644.9 m against 0.487 m for the constant one: the filter loses the solution. The gate rejects 40699 features there, thirteen times more than in the neighboring cells. With the additive convention the same configuration survives (0.383 m): the divergence is not a property of the threshold itself, it arises in the interaction.

The divergence is removed by matching the noise scale, not by changing the form of the rule. The cause is not that the threshold is too strict, but that it is strict *for the adopted* σ . At $\sigma = 3$ px the same quantile threshold gives 0.436 m on the same cell, rejecting 2504 features instead of 40699, and at the same time it *outperforms the constant threshold at the same* σ (0.511 m). The gain therefore does not reduce to weakening the rejection by increasing σ : the quantile form is genuinely more accurate when the noise scale is matched to it. An alternative explanation, that the divergence is seeded during the unsettled start, was tested and not confirmed: if the gate is not switched on for the first 20 ... 200 updates, the error does not fall but grows (656–1785 m), and the rejection count does not decrease but rises from 40699 to 56035.

No single noise scale serves the grid. To find out whether the interaction observed on one cell extends to the rest, the whole grid was repeated at $\sigma = 3$ px with the quantile threshold (Table 4). It does extend, but not in the direction one might expect from the diverging cell alone: raising σ repairs ORB+KLT and degrades everything else, by 14.3 to 55.2 %.

Table 4.

Mean ATE3D (m) of the right-invariant convention at two noise scales				
front end	constant, $\sigma = 1.5$	quantile, $\sigma = 1.5$	quantile, $\sigma = 3$	change
XFeat-1024 + KLT	0.337	0.266	0.304	+14.3 %
ORB-1024 + KLT*	0.356	0.284	0.341	+20.2 %
XFeat-1024 + desc	0.177	0.136	0.181	+33.5 %
ORB-1024 + desc	0.180	0.179	0.278	+55.2 %

* without MH01, so that the two scales are compared on the same four sequences. On MH01 itself the ordering is the opposite: 644.9 m at $\sigma = 1.5$ against 0.436 m at $\sigma = 3$. The disagreement is therefore not between a front end and a scale but between a *sequence* and a scale: within one front end, one sequence requires $\sigma = 3$ while the remaining four are 9 to 39 % worse at that value.

Hence the main practical statement of the work: *the form of the threshold and the adopted measurement-noise scale are not amenable to separate tuning*. Transferring one component without the other, that is, taking the classical criterion while keeping σ from the previous setting, costs the solution; but neither is there a single scale that suits the whole grid, so the pair has to be chosen per configuration rather than once for the method.

The contribution of the convention is smaller and does not transfer. The right convention is better than the additive one by 4.1, 8.2, 2.4 and 10.0 % on average with the constant threshold and by 3.7, 4.9, 7.3 and 9.9 % with the quantile one, that is, an order of magnitude less than the choice of the threshold form gives, with one exception: on ORB+desc the ratio *reverses*, the form of the threshold gives nothing while the convention gives almost ten per cent. The claim that “the measurement acceptance rule matters more than the definition of the error” holds for three configurations out of four and is not general.

Even the division of configurations into stable and unstable does not transfer. The per-cell ratio ATE_{add}/ATE_{right} shows in how many cells the advantage of the right convention is retained; the range of this ratio is given in parentheses (Table 5).

Table 5.

Stability of the advantage of the right-invariant convention under two threshold forms

front end	constant 5.99m	quantile $\chi^2_{m,0.95}$
XFeat-1024 + KLT	5/5 (1.012–1.146)	4/5 (0.954–1.144)
ORB-1024 + KLT	4/5 (0.950–1.294)	4/4 (1.021–1.176)
XFeat-1024 + desc	3/5 (0.919–1.177)	4/5 (0.996–1.187)
ORB-1024 + desc	5/5 (1.022–1.183)	5/5 (1.029–1.191)

With the constant threshold the stable and unstable configurations are distributed *crosswise*: XFeat + KLT and ORB + desc are stable, ORB + KLT and XFeat + desc are not, so neither detector nor tracking method is stable on its own: it is the pair that decides. With the quantile threshold this division is rebuilt and starts to be determined by the *detector* itself: both configurations with ORB are stable, both with XFeat are not. Thus not only the magnitude of the effect changes, but the very structure of the dependence, as soon as the form of the rule is changed. A control run with Shi-Tomasi corners and optical-flow tracking gives 0.938–1.124; this configuration does not enter the 2×2 design, because corners have no descriptor.

Limits of what was obtained. What has been measured concerns five sequences of a single dataset and four front ends, and the comparison is made by position accuracy only: these measurements say nothing about the consistency of the estimate, where the advantage of invariant conventions shows more strongly. The noise scale was varied at two values only, 1.5 and 3 px; a denser sweep could locate the scale that suits each configuration, which was not the aim here. Graduated forms of weighting, in which rejection is replaced by a smooth reduction of trust, are not considered here: they constitute a separate subject.

Conclusions

A mathematical model of the filter on the group $SE_2(3)$ is presented in a completeness sufficient for reproduction; only the navigation block of the transition matrix is independent of the estimate, whereas the columns associated with the IMU biases are not.

The widely used constant rejection threshold does not agree with the dimension of the measurement: 5.99 is the quantile $\chi^2_{2,0.95}$, a value for a two-component measurement, whereas in the stereo setting with landmark elimination the residual dimension is $4M - 3 \geq 5$. It is softer than the quantile one by a factor of 2.7–4.4 and non-uniform in track length. Replacing it by the quantile threshold reduces the mean error by 21.1, 23.4 and 20.2 % on three front ends and gives nothing on the fourth: the form of the threshold matters more than the presence of the gate and an order of magnitude more than the error convention (2.4–10.0 %).

The threshold form and the noise scale are not separable, and no single scale serves the grid: at $\sigma = 1.5$ px one configuration loses the solution (644.9 m), at $\sigma = 3$ px the same form gives 0.436 m, yet that scale costs 14.3–55.2 % on the other front ends and 9–39 % on the remaining sequences of the same one. The pair therefore has to be chosen per configuration rather than once for the method.

The limits of transferability are twofold: the ratio of the two contributions reverses on ORB + desc, and the division of configurations into stable and unstable is rebuilt when the form of the rule changes, passing from the pair “detector + tracking” to the detector alone. Further research should measure the same factors by consistency indicators rather than by ATE alone.

ADDITIONAL INFORMATION

AUTHOR CONTRIBUTIONS

Conceptualisation, R.M.; methodology, A.Zh. and R.M.; software, A.Zh.; validation, A.Zh.; formal analysis, A.Zh.; investigation, A.Zh.; data curation, A.Zh.; writing—original draft preparation, A.Zh.; writing—review and editing, R.M.; visualisation, A.Zh.; supervision, R.M. All authors have read and agreed to the published version of the manuscript.

DECLARATION ON THE USE OF GENERATIVE ARTIFICIAL INTELLIGENCE TOOLS

During the preparation of this work, the authors used Claude 5 Opus in order to perform grammar and spelling checks, translate specific phrases into English, and verify mathematical formulas. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication’s content.

REFERENCES

1. Cadena, C., Carlone, L., Carrillo, H., Latif, Y., Scaramuzza, D., Neira, J., Reid, I., Leonard, J.J.: Past, present, and future of simultaneous localization and mapping: toward the robust-perception age. *IEEE Transactions on Robotics* 32(6), 1309–1332 (2016)

2. Huang, G.: Visual-inertial navigation: a concise review. In: IEEE International Conference on Robotics and Automation (ICRA), pp. 9572–9582 (2019)
3. Mourikis, A.I., Roumeliotis, S.I.: A multi-state constraint Kalman filter for vision-aided inertial navigation. In: IEEE International Conference on Robotics and Automation (ICRA), pp. 3565–3572 (2007)
4. Barrau, A., Bonnabel, S.: The invariant extended Kalman filter as a stable observer. IEEE Transactions on Automatic Control 62(4), 1797–1812 (2017)
5. Geneva, P., Ekenhoff, K., Lee, W., Yang, Y., Huang, G.: OpenVINS: a research platform for visual-inertial estimation. In: IEEE International Conference on Robotics and Automation (ICRA), pp. 4666–4672 (2020)
6. Sun, K., Mohta, K., Pfrommer, B., Watterson, M., Liu, S., Mulgaonkar, Y., Taylor, C.J., Kumar, V.: Robust stereo visual inertial odometry for fast autonomous flight. IEEE Robotics and Automation Letters 3(2), 965–972 (2018)
7. Forster, C., Carlone, L., Dellaert, F., Scaramuzza, D.: On-manifold preintegration for real-time visual-inertial odometry. IEEE Transactions on Robotics 33(1), 1–21 (2017)
8. Fornasier, A., van Goor, P., Allak, E., Mahony, R., Weiss, S.: MSCEqF: a multi state constraint equivariant filter for vision-aided inertial navigation. IEEE Robotics and Automation Letters (2024). arXiv:2311.11649
9. Hao, N., Tian, C., He, F.: A transformation-based consistent estimation framework: analysis, design and applications. arXiv:2502.05008 (2025)
10. Ge, Y., Delama, G., Scheiber, M., Fornasier, A., van Goor, P., Weiss, S., Mahony, R.: The difference between the left and right invariant extended Kalman filter. arXiv:2507.04568 (2025)
11. Yang, Y., Chen, C., Lee, W., Huang, G.: Decoupled right invariant error states for consistent visual-inertial navigation. IEEE Robotics and Automation Letters 7(2), 1627–1634 (2022)
12. Жарков, А., Маслій, Р., Гармаш, В.: Аналіз підходів Visual SLAM для задачі навігації автономного робота. Вісник Хмельницького національного університету, 2024, № 3, т. 1 (335), с. 67–77. <https://doi.org/10.31891/2307-5732-2024-335-3-10>
13. Жарков, А.В., Маслій, Р.В.: Візуально-інерційний SLAM з використанням розширеного фільтра Калмана для автономної навігації. Вісник Вінницького політехнічного інституту, 2025, № 2 (179), с. 118–126. <https://doi.org/10.31649/1997-9266-2025-179-2-118-126>
14. Burri, M., Nikolic, J., Gohl, P., Schneider, T., Rehder, J., Omari, S., Achtelik, M.W., Siegwart, R.: The EuRoC micro aerial vehicle datasets. The International Journal of Robotics Research 35(10), 1157–1163 (2016)
15. Potje, G., Cadar, F., Araujo, A., Martins, R., Nascimento, E.R.: XFeat: accelerated features for lightweight image matching. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2682–2691 (2024)
16. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: ORB: an efficient alternative to SIFT or SURF. In: IEEE International Conference on Computer Vision (ICCV), pp. 2564–2571 (2011).
17. Shi, J., Tomasi, C.: Good features to track. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 593–600 (1994)

Анатолій ЖАРКОВ, Роман МАСЛІЙ
 Вінницький національний технічний університет

ПРАВОІНВАРІАНТНА СТЕРЕОВІЗУАЛЬНО-ІНЕРЦІАЛЬНА ОДОМЕТРІЯ: ПОРОГИ ВІДБРАКОВУВАННЯ ТА МАСШТАБ ШУМУ

Предметом статті є фільтрове візуально-інерціальне оцінювання стану безпілотного літального апарата за показами інерціального вимірювального пристрою та стереозображень. Викладено математичну модель фільтра на матричній групі $LiSE_2(3)$ з приєднаними зміщеннями, з виключенням орієнтирів проекцією нев'язки на нуль-простір якобіана та з відбраковуванням за нормованою відстанню Махаланобіса, у двох конвенціях похибки: правоінваріантній та класичній адитивній.

Експерименти проведено на п'яти послідовностях EuRoC MH із чотирма фронтендами за схемою 2×2 : два способи ведення точки між кадрами на двох детекторах, навченому XFeat та класичному ORB. Показано, що поширений сталий поріг відповідає квантилю розподілу χ^2 для двокомпонентного вимірювання; заміна його квантильним порогом, узгодженим із фактичною розмірністю нев'язки, зменшує середню похибку положення на 20–23 % на трьох фронтендах із чотирьох, тоді як внесок конвенції похибки становить 2,4–10,0 %.

Наукова новизна полягає у встановленні того, що форма порога відбраковування та прийнятий масштаб шуму вимірювань підлягають спільному, а не роздільному калібруванню: поріг, визначений квантилем χ^2 , не потребує підбраної сталої, але перенесений без узгодження σ коштує розв'язку. Практична значущість полягає у призначенні порога без ручного добору та в кількісно окреслених межах переносимості налаштувань між фронтендами.

Ключові слова: візуально-інерціальна одометрія; інваріантний фільтр Калмана; група Li ; безпілотний літальний апарат; відбраковування викидів; відстань Махаланобіса; стереозір; EuRoC.