Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Gyroscope-Guided Optimal Transport for SuperGlue: A Training-Free Epipolar-Arc Prior

Ufuk Asil

ufuk.asil@ostimteknik.edu.tr

Ostim Technical University

Research Article

Keywords: Feature matching, SuperGlue, Optimal transport, Gyroscope, Visual-inertial odometry, Epipolar geometry, Training-free

Posted Date: November 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11349008/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Gyroscope-Guided Optimal Transport for SuperGlue: A Training-Free Epipolar-Arc Prior

Ufuk Asil[1*]

[1*]OSTIM Technical University, 100. Yil Bulvari 55/F, OSTIM, Yenimahalle, Ankara, 06374, Turkey.

Corresponding author(s). E-mail(s): ufuk.asil@ostimteknik.edu.tr;

Abstract

Learned matchers such as SuperGlue use image evidence only, whereas visual–inertial platforms carry a gyroscope that measures the inter-frame rotation independently of image quality. Added to a pretrained SuperGlue as a log-prior around the pure-rotation prediction, this rotation did not improve pose accuracy: translation-induced parallax forces a loose tolerance, and because the dustbin receives no prior, a penalty-only prior shifts mass towards "unmatched". We therefore propose a training-free epipolar-arc prior (EAP). With the gyroscope rotation fixed, a robust two-point solver estimates the translation direction from the same network pass, and candidates are rewarded by their angular distance to the great-circle arc on which a correct match must lie. With hyperparameters frozen on one TUM-VI sequence, the EAP raises the area under the pose-error curve at **10°** (AUC@**10°**) of a 5-point pose pipeline on 15 held-out TUM-VI and EuRoC sequences (200–600 ms) by +1.9 and +2.2 points on unmodified images and by up to +9.9 points under synthetic motion blur; the gain is positive on all 15 sequences and grows with rotation. Over all regimes it gains +4.4 and +2.7 points, versus +1.0 and +0.1 for a post-hoc arc filter and +1.5 and +0.5 for a gyroscope-free arc. Transferred unchanged to LightGlue, it gains +2.3 and +0.8 points, significantly only under motion blur and, on TUM-VI, a low keypoint budget. With a pose solver that uses the gyroscope rotation, it adds only +0.5 and +0.2 points, so the EAP mainly benefits pipelines whose pose solver ignores inertial data.

Keywords: Feature matching, SuperGlue, Optimal transport, Gyroscope, Visual-inertial odometry, Epipolar geometry, Training-free

1 Introduction

Feature matching between consecutive frames forms the front end of visual odometry (VO), visual–inertial odometry (VIO) and simultaneous localization and mapping (SLAM): the correspondences it produces feed relative-pose estimation, triangulation and bundle adjustment. Learned matchers such as SuperGlue [1] and LightGlue [2], built on learned keypoints and descriptors [3], are now widely used for this task, yet they rely on image evidence alone. Fast camera motion causes large, spatially varying displacements between frames and changes the keypoint positions that the matcher's positional encoding (PE) sees; it is also often accompanied by blur, which degrades the descriptors. Knowing where a correct match can lie is therefore valuable, even though our measurements show that, in the sequences we study, SuperGlue's pose accuracy does not decrease with the rotation angle as such.

Platforms that run VIO carry a gyroscope. Over the short interval between two frames, the integrated angular rate predicts the camera rotation independently of image quality, with an error below about one degree in our data (Section 5.1). Classical trackers have long exploited this, warping or pre-positioning the search window with the gyroscope rotation [4, 5], and known-rotation solvers reduce relative pose to a two-point problem for the translation direction [6, 7]. Modern learned matchers, by contrast, usually run as visual-only black boxes; inertial data enter the pipeline only after matching. End-to-end inertial–visual networks do exist [8, 9], but they are trained on paired gyroscope and image data and are not designed to be attached to an existing pretrained matcher.

This paper asks how the gyroscope can be used *inside* a pretrained learned matcher without retraining it. SuperGlue is suited to this question because its final stage solves an explicit optimal transport (OT) problem on a score matrix, to which geometric knowledge can be added as a log-potential. The most direct option—penalizing candidates far from the pure-rotation (infinite-homography) prediction—was ineffective in our experiments: the rotation-only priors we evaluated changed relative-pose accuracy by at most a few tenths of a point, mostly downwards. Our analysis attributes this to two causes. First, at the time baselines relevant for keyframe-based VO, translation-induced parallax amounts to several degrees, so a rotation-only prior must be either too loose to help or too tight to be correct. Second, SuperGlue's dustbin does not receive the prior, so a penalty-only prior shifts probability mass towards "unmatched", and the Sinkhorn normalization does not compensate for this shift.

We address both causes. Parallax is not isotropic: in a static scene, the correct match must lie on a great-circle arc that runs from the rotated bearing (infinite depth) towards the translation direction (zero depth). The gyroscope rotation fixes the start of this arc, while the translation direction can be estimated from the same graph neural network (GNN) pass with a known-rotation two-point solver. The resulting *epipolar-arc prior* (EAP) is tight across the arc and free along it. We inject it with a reward-only offset, which raises the scores of consistent candidates and leaves those of inconsistent candidates unchanged. In addition, we evaluate rotation compensation of the PE and image warping by the gyroscope rotation.

Our contributions are:

- A training-free epipolar-arc log-prior for the OT layer of a pretrained SuperGlue, parameterized by the gyroscope rotation and by a translation direction estimated from the same GNN pass. It requires no second GNN pass, but does require one additional Sinkhorn solve and a two-point random sample consensus (RANSAC) step, which together take 23 ms per pair in our unoptimized implementation (Section 3.5). Applied to LightGlue without re-tuning, it yields smaller gains, which are significant only under synthetic motion blur and, on TUM-VI, with a low keypoint budget (Section 5.6).
- An analysis of the role of the dustbin in prior injection, showing that the offset of the prior is not a gauge and that penalty-only priors reduce matches; a development-set study of the offset supports this analysis (Sections 3.3 and 5.2).
- A held-out evaluation protocol for this question. Hyperparameters are selected on one development sequence and frozen, after which the methods are evaluated on 15 unseen sequences from two datasets and two camera models, covering four visual regimes and four time baselines. RANSAC variability is controlled by point-order permutations (in OpenCV 5.0, setting the random seed did not change the RANSAC result in our test, whereas the point order did); pre-declared hypotheses are tested with Holm-corrected paired permutation tests; and dependence-robust sequence-level analyses are reported (Sections 4–5).
- A separation of where the gyroscope helps: inside the matcher (more epipolar-consistent correspondences, better pose with a standard 5-point solver) versus inside the pose solver (known-rotation 2-point solver), together with controls that apply the same geometry post hoc, estimate it without the gyroscope, or change the match/no-match balance without geometry (Section 5).

Section 2 reviews related work and Section 3 presents the method. Section 4 describes the evaluation protocol, Section 5 reports the results, Section 6 discusses limitations, and Section 7 concludes.

2 Related Work

2.1 Learned Feature Matching and Optimal Transport

SuperGlue [1] matches SuperPoint [3] keypoints with an attention-based GNN followed by an entropy-regularized OT layer solved with Sinkhorn iterations [10, 11]; unmatched points are absorbed by a dustbin. LightGlue [2] replaces the OT layer by a lighter assignment head with adaptive depth, and detector-free matchers such as LoFTR [12] operate on dense feature maps. Epipolar geometry has also been built into trained matchers, for example through epipolar attention [13], as a training-time loss [14], or by feeding noisy 3D cues to the GNN [15]. Each of these requires training. We build on SuperGlue because our intervention needs an explicit score matrix that is subsequently normalized into a partial assignment, and SuperGlue exposes this interface directly. Outside learned matching, geometry has also served as an OT cost in classical, training-free settings: HOT-POT [16] formulates sparse stereo matching as (partial) OT with an epipolar cost. In contrast, we keep the learned scores of a pretrained matcher and add a gyroscope-derived epipolar term to them.

2.2 Rotation Robustness of Learned Matchers

SuperPoint descriptors and the PE of SuperGlue are not rotation invariant. One line of work makes descriptors rotation equivariant by design or by learning; Steerers [17], for example, learn a linear map in descriptor space that approximates image rotation. Such approaches require (re)training, and they address in-plane rotation of the image content. Our PE compensation complements them: it is training-free, uses a measured 3D rotation, and acts only on positions.

2.3 Inertial Aiding of Visual Correspondence

Using gyroscope measurements to predict feature motion is well established. Hwangbo et al. [4, 5] used the integrated rotation to initialize the warp parameters of a Kanade–Lucas–Tomasi (KLT) tracker [18, 19], and later work extended this idea with target-free camera–gyroscope calibration [20] and robustness to large motion [21]. Visual SLAM systems guide feature search with motion priors, for example the constant-velocity model of ORB-SLAM [22]. In VIO, inertial measurement unit (IMU) measurements are integrated between frames [23] and fused with visual constraints in the estimator [24–26]; the predicted motion can likewise be used to restrict the search for correspondences. End-to-end approaches such as GyroFlow [8, 9] fuse gyroscope data into deep optical-flow networks and require training. Closest to our setting, Zhao et al. [27] inject an IMU-derived spatial prior as a probabilistic attention term into an attentional GNN keypoint matcher, and Lu et al. [28] feed epipolar lines from a regressed coarse pose as additional features to a SuperGlue-style matcher; both retrain the matcher, and the latter uses an image-derived rather than inertial pose. Learned front ends have also been combined with VIO back ends without inertial guidance of the matching itself [29]. Our method instead adds inertial geometry to the scores of a pretrained matcher, without retraining or changing its weights.

2.4 Known-Rotation Relative Pose

When the relative rotation is known, the essential-matrix constraint becomes linear in the translation direction and two correspondences suffice [6, 7]; partially known orientation similarly reduces the minimal problem [30]. Such solvers are used for efficient outlier rejection in camera–IMU systems. We use the 2-point solver in two roles: inside the matcher, to estimate the translation direction that parameterizes the EAP, and as a downstream pose solver in a separate analysis. This second role lets us separate the effect of the matcher's correspondences from that of using the gyroscope rotation in the solver.

2.5 Robust Estimation and Guided Matching

RANSAC [31] and its variants [32, 33] remove outliers after matching, and guided matching restricts the search to a neighbourhood of an epipolar line once an initial geometry is available (34, Ch. 11; 35). Our EAP borrows the guided-matching idea, but it acts as a soft, truncated term inside the assignment problem rather than as

a hard search window or a post-hoc filter. We compare both placements directly (Section 3.7).

3 Method

Our method keeps the SuperPoint detector, the SuperGlue GNN and their pretrained weights fixed. Only two inputs of the pretrained matcher change: the score matrix that enters the Sinkhorn assignment layer (Sections 3.3–3.5) and, optionally, the keypoint positions that enter the positional encoder (Section 3.6). As Figure 1 shows, the pretrained GNN runs once, and the Sinkhorn layer assigns its score matrix S twice. Stage 1 produces provisional matches; from these, the translation direction $\hat{t}$ is estimated while the gyroscope rotation R is held fixed. Stage 2 then re-solves the assignment after the epipolar-arc log-prior $P(R, \hat{t})$ has been added to S. The optional PE compensation instead feeds rotation-compensated keypoint positions to the positional encoder, so its GNN pass cannot be shared with the uncompensated matcher.

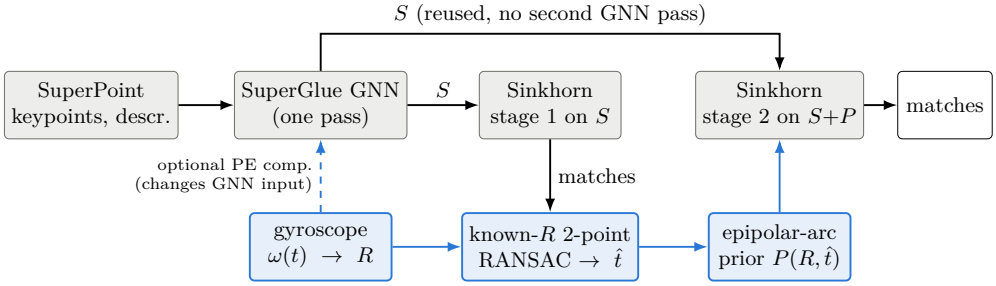


Fig. 1 Overview of the gyroscope-guided matcher: one GNN pass, two Sinkhorn solves, and the epipolar-arc prior $P(R, \hat{t})$ added to the scores in stage 2; the dashed arrow marks the optional positional-encoding (PE) compensation.

3.1 Notation and Camera Model

Let I_a and I_b be two frames of a calibrated camera taken at $t_a < t_b$. SuperPoint [3] yields keypoints x_i^a $(i = 1, \ldots, M)$ and x_j^b $(j = 1, \ldots, N)$. For every keypoint we use the unit bearing vector $f = \pi^{-1}(x)/\|\pi^{-1}(x)\|$, where π^{-1} is the unprojection of the respective camera model (Kannala–Brandt fisheye [36] for TUM-VI, pinhole with radial–tangential distortion for EuRoC MAV and the Hwangbo sequences). All priors below are defined on bearing vectors, i.e. in angular units on the unit sphere, and are therefore independent of the lens model. Pixel-domain priors, by contrast, are affected by the anisotropic stretching that pixel distances undergo at the periphery of a fisheye image after undistortion; the angular formulation avoids it. For the fisheye model we obtain the bearing by inverting the Kannala–Brandt polynomial directly, so that rays beyond 90° from the optical axis (about 8% of the TUM-VI image area) are represented correctly.

The relative motion is written as $f^b \propto R\,X^a + t$ for a scene point X^a expressed in the frame of camera a, with $R \in \mathrm{SO}(3)$ and translation t (known only up to scale).

3.2 Gyroscope Integration

The gyroscope measures $\omega_{\mathrm{m}}(t) = \omega(t) + b_g + \eta_g(t)$ in its own frame, and we map this rate to the camera frame with the camera–IMU extrinsic rotation (or, for the Hwangbo data, with the supplied calibration matrix C_g). When a standstill window exists (a 1 s window with per-axis standard deviation below 0.02 rad/s), the bias b_g is estimated as the mean angular rate over it. Otherwise the bias is not removed, and its effect is contained in the measured integration error that we report. We obtain the relative rotation by first-order quaternion integration with $\delta t = 2$ ms and linear interpolation of the angular rate,

$$q_{k+1} = \mathrm{norm}\Big(q_k + \tfrac{1}{2}\,\Omega\big(\omega(\tau_k)\big)\,q_k\,\delta t\Big), \tag{1}$$

which yields the rotation R_{ab} of camera b relative to camera a. Points are transferred with $R = R_{ab}^{\top}$, and we verified the direction of this transfer empirically (Section 5.1).

3.3 Log-Prior Injection into the Assignment Layer

After its GNN, SuperGlue computes descriptor scores S_{ij}, augments the $M \times N$ matrix with a dustbin row and column that carry a single learned score, and solves an entropy-regularized partial assignment with Sinkhorn iterations [1, 10]. We add a geometric log-potential to the real entries only:

$$S'_{ij} = S_{ij} + \alpha\Big(\max\big(-\tfrac{1}{2}q_{ij},\,-\kappa\big) + s\Big), \tag{2}$$

where $q_{ij} \geq 0$ is a squared, normalized geometric discrepancy between keypoint i and candidate j (defined below), κ truncates the penalty so that geometrically inconsistent candidates are not excluded outright, α is a weight (we use $\alpha = 1$) and $s \in [0, \kappa]$ is an offset. No prior is applied to keypoints whose gyroscope prediction is undefined, i.e. whose predicted ray falls behind the camera.

The offset is not a gauge.

In a *square* transport problem, a constant added to all entries of a row is absorbed by the Sinkhorn dual variables and leaves the assignment unchanged. That argument does not carry over here, because the dustbin column does not receive the prior: adding a constant c to the real entries of row i multiplies their transport kernel by e^c *relative to the dustbin*, and therefore shifts mass between "matched" and "unmatched". A penalty-only prior ($s = 0$, q-dependent term in $[-\kappa, 0]$) consequently pushes keypoints towards the dustbin and reduces the number of matches. A reward-only prior ($s = \kappa$), in contrast, leaves the scores of geometrically inconsistent candidates unchanged and raises those of consistent candidates relative to the dustbin. (The assignment of inconsistent candidates can still change, because Sinkhorn normalizes rows and columns

jointly.) Equivalently, a reward-only prior is a penalty-only prior combined with lowering the dustbin score by κ; net of both, only candidates inside the arc band gain relative to the dustbin. Since the reward-only prior also increases the number of matches, Section 5.5 tests whether a geometry-free shift of the match/no-match balance explains the gain. By the same argument, a per-row normalizer such as $-\frac{1}{2}\log|\Sigma_i|$ in a Mahalanobis prior is *not* compensated by Sinkhorn. We therefore treat s as a hyperparameter and select it on the development sequence (Section 4.6).

Equation (2) can be read, by analogy, as adding the log of a robust (truncated Gaussian, i.e. Gaussian-plus-uniform) geometric likelihood to the matching score, in the spirit of mixture models such as MLESAC [32]. The analogy is only heuristic. The GNN scores are not calibrated log-likelihoods, and the geometric and visual evidence are not conditionally independent; for this reason, the weight α and offset s are determined empirically rather than derived.

3.4 Rotation-Only Prior

Under pure rotation (or for distant points) the bearing of x_i^a in camera b is Rf_i^a. The rotation-only prior uses

$$q_{ij}^{\mathrm{rot}} = \frac{\angle\big(Rf_i^a,\, f_j^b\big)^2}{\sigma_R^2}, \tag{3}$$

with an angular tolerance σ_R that must absorb gyroscope error, time-synchronization error and the translation-induced parallax. Parallax grows with the time baseline and as the scene comes closer; in the hand-held indoor sequences considered here it is not negligible and can reach several degrees at a 600 ms baseline. A rotation-only prior must therefore trade tightness against tolerance to parallax. The pixel-domain isotropic and Mahalanobis priors belong to this class and are kept as baselines (`iso`, `maha`): the former is an isotropic truncated Gaussian around the infinite-homography prediction $KR^\top K^{-1}x$, and the latter its anisotropic variant with propagated rotation, synchronization and isotropic parallax covariance.

3.5 Epipolar-Arc Prior

Parallax is not isotropic: for a static scene, $f_j^b \propto z\,Rf_i^a + t$ with depth $z > 0$. Hence the correct match lies in the epipolar plane spanned by Rf_i^a and t, and more specifically on the great-circle *arc* that starts at Rf_i^a (infinite depth) and ends at $\hat{t} = t/\|t\|$ (zero depth). The arc thus encodes not only the epipolar line but also the direction in which parallax displaces the point, leaving only the unknown depth free. We define

$$q_{ij}^{\mathrm{arc}} = \frac{d_{\mathrm{arc}}\big(f_j^b;\, Rf_i^a, \hat{t}\big)^2}{\sigma_E^2}, \tag{4}$$

where d_{arc} is the angular distance of f_j^b to the arc: with the plane normal $n_i = Rf_i^a \times \hat{t}/\|Rf_i^a \times \hat{t}\|$, it equals $\arcsin|f_j^{b\top} n_i|$ if the projection of f_j^b onto the plane falls between the arc's end points, and the angle to the nearer end point otherwise. Because depth is left free, σ_E only has to absorb rotation and translation-direction errors and

can be much tighter than σ_R. Fig. 2 illustrates this on a pair of the development sequence. We drew 14 keypoints at random among the 40 ground-truth-consistent SuperGlue matches with the largest residual to the rotation-only prediction; their matched points lie tens of pixels away from the rotation-only predictions but close to the arcs, which were computed with a $\hat{t}$ estimated from the stage-1 matches (angular error 1.3°).

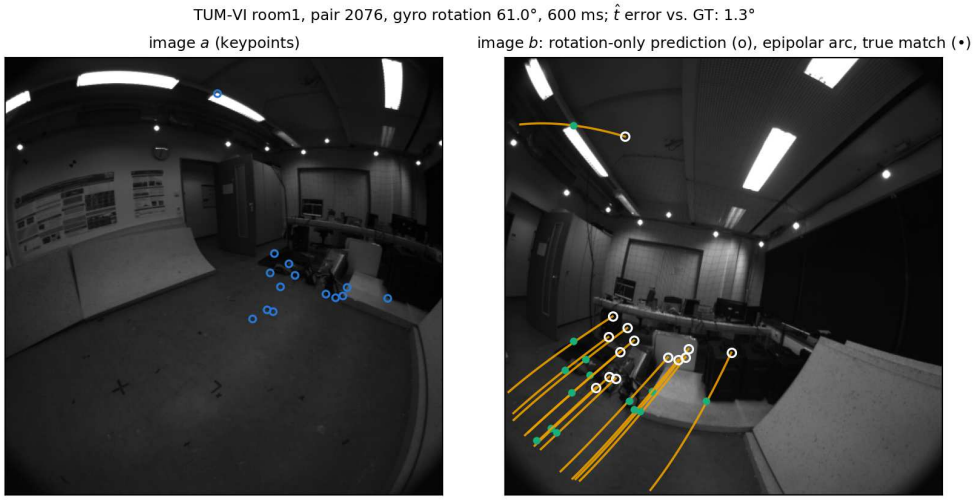


Fig. 2 Epipolar arcs on a development-sequence pair (TUM-VI `room1`, 61° gyroscope rotation, 600 ms): keypoints in image a (left) and, in image b, their rotation-only predictions (white circles), the arcs towards $\hat{t}$ (yellow) and the matched points (green).

Estimating the translation direction.

The gyroscope does not measure $\hat{t}$, so we recover it from the same GNN pass: the score matrix S is first assigned without prior (stage 1, identical to the original matcher), and $\hat{t}$ is estimated from the stage-1 matches with R held fixed. Known rotation makes the epipolar constraint linear in t, $t^\top(Rf_k^a \times f_k^b) = 0$, so two correspondences determine $\hat{t}$ up to sign [6, 7]. We run a 2-point RANSAC [31] (256 hypotheses, Sampson threshold of 2 px/f_x), refine $\hat{t}$ by singular value decomposition (SVD) on the inliers, and resolve the sign by cheirality. When the pure-rotation model explains at least 90% as many stage-1 matches as the estimated epipolar geometry, we consider the translation unobservable and use the rotation-only prior (Eq. 3) instead. Stage 2 then solves the assignment on $S + P^{\text{arc}}$. Since the GNN is not re-run, the additional cost amounts to one Sinkhorn solve and the 2-point RANSAC.

The final matches are an output of the matcher and can be consumed downstream by any pose solver. To keep improvements of the matcher from being confounded with the choice of solver, our pose evaluation deliberately uses the standard rotation-agnostic 5-point solver [37] for all methods; the known-rotation 2-point solver is reported separately (Section 5.8).

3.6 Rotation-Compensated Positional Encoding

SuperGlue adds an encoding of the keypoint position to each descriptor before the GNN. Because large rotations can shift the image positions of corresponding points by hundreds of pixels, the positional cues seen by the attention layers disagree between the two images. As an optional component (PE compensation), we therefore feed the positional encoder of image a with the rotation-compensated positions $\pi(Rf_i^a)$ instead of x_i^a and leave the descriptors unchanged. To keep the encoder input within its training range, positions that would fall behind the camera or more than 25% of the image size outside the image are left unchanged. For the radial–tangential model, rays beyond the calibrated field of view are projected without distortion, because the distortion polynomial folds over outside that region. PE compensation changes the GNN input, so its GNN pass cannot be shared with the original matcher. It can be combined with the EAP, in which case stage 1 uses the compensated GNN pass. Unlike rotation-equivariant descriptors such as Steerers [17], this component acts only on positions and requires no training.

3.7 Control: Post-Hoc Gyroscope Gate

To test whether injecting geometry *inside* the assignment matters, we also evaluate a post-hoc gate that applies the same geometry *after* matching: original SuperGlue matches with $q_{ij}^{\mathrm{arc}} > 2\kappa$ (i.e. outside the truncation radius of the prior) are discarded. The gate shares $\hat{t}$, σ_E and κ with the EAP.

3.8 Application to LightGlue

LightGlue [2] replaces the dustbin-augmented Sinkhorn layer by a double softmax over a similarity matrix S, combined with per-point matchability scores z: $\log P_{ij} = \log \mathrm{softmax}_j(S_{ij}) + \log \mathrm{softmax}_i(S_{ij}) + \log \sigma(z_i^a) + \log \sigma(z_j^b)$. We add the same prior P to S after the last transformer layer and before the double softmax; stage 1 is the unmodified LightGlue assignment, and $\hat{t}$ is estimated as in Section 3.5. This structure leads to two differences from SuperGlue. First, the double softmax is invariant to constants added to a row or a column, so the offset s has no effect and the prior can only redistribute probability among the candidates of a keypoint. Second, because the prior leaves the matchability scores untouched, it does not shift mass between "matched" and "unmatched". We disable LightGlue's adaptive depth and point pruning so that all layers are evaluated, and we reuse the hyperparameters selected for SuperGlue without re-tuning.

4 Experimental Setup

4.1 Datasets and Development/Test Split

We use three public datasets that provide synchronized camera and gyroscope data (Table 1). The EuRoC `MH03` sequence was excluded because most images of our local copy are unreadable. All hyperparameters were selected on a single development sequence, TUM-VI `room1`, and frozen before any held-out sequence was processed.

Table 1 Development/test split: sequences, camera model and resolution, and ground-truth source (GT: ground-truth relative pose).

Role	Sequences	Camera, resolution [px]	GT
Development	TUM-VI `room1` [38]	fisheye, 512^2	motion capture
Held-out test	TUM-VI `room2`–`room6`	fisheye, 512^2	motion capture
Held-out test	EuRoC `MH01,02,04,05`, `V1_01--03`, `V2_01--03` [39]	pinhole, 752×480	Leica / Vicon
Consistency only	Hwangbo `desk`, `aerial` [4, 5]	pinhole, 640×480 / 320×240	none

The selection criterion, the held-out sequences and the list of hypotheses to be tested were recorded beforehand in an internal (not externally registered) protocol file. This file is part of the code release, together with a time-stamped log of all deviations; entries added or corrected after our internal review are marked as such (Section 4.5).

TUM-VI `room` sequences are hand-held recordings with fast rotations in a motion-capture room, whereas EuRoC contains micro aerial vehicle flights in an industrial hall (`MH`) and in a motion-capture room (`V`). Both datasets capture frames at 20 Hz and IMU measurements at 200 Hz. For EuRoC we rely on the provided ground-truth state estimates. Because these estimates fuse motion capture or laser tracking with the IMU, their orientation is itself an estimate and is not independent of the gyroscope; the gyroscope errors we report for EuRoC (and the known-rotation solver results) may therefore be optimistic.

4.2 Pairs and Visual Regimes

For every sequence we form 120 uniformly spaced pairs $(I_t, I_{t+\text{skip}})$ for each skip $\in \{4, 8, 12\}$, i.e. time baselines of 200, 400 and 600 ms. Each pair is evaluated in four visual regimes: *full* (up to 512 keypoints), *low budget* (up to 128 keypoints), *Gaussian blur* (9×9 kernel, $\sigma \approx 1.7$ px, the OpenCV default for this size) and *synthetic motion blur* (a 15-pixel linear kernel with a random orientation drawn per frame from a fixed seed). Both blur regimes are synthetic image degradations rather than physically rendered motion blur. In total, the held-out set contains $15 \times 3 \times 4 \times 120 = 21{,}600$ pair evaluations per method.

4.3 Metrics

Relative pose.

Matches are first converted to normalized bearing coordinates. The essential matrix is then estimated with OpenCV's 5-point solver [37] inside RANSAC [31] (threshold 1 px/f_x, confidence 0.9999) and decomposed with `recoverPose`. We define the pose error as $\max(\text{err}_R, \text{err}_t)$, where err_t is the angle between translation directions, and report the area under its cumulative curve (AUC) up to 5°/10°/20°, following Sarlin et al. [1].

RANSAC variability.

In our tests with OpenCV 5.0, the output of `findEssentialMat` with RANSAC was insensitive to `cv2.setRNGSeed` but sensitive to the order of the input points. On

five pairs of the development sequence, 30 different seeds produced identical results, whereas 30 random permutations of the same matches produced pose errors that varied by up to 6.3° for a single pair. To account for this order dependence, we repeat every pose estimate over random point-order permutations ($P = 10$ on the development sequence, $P = 5$ on held-out data) and report the AUC averaged over permutations together with its standard deviation across permutations.

Epipolar consistency of matches.

A match counts as epipolar-consistent if its Sampson distance [34, Ch. 11] to the ground-truth epipolar geometry is below 3 px/f_x. The measure is deterministic and does not involve RANSAC. It cannot, however, detect wrong matches that lie on the correct epipolar line, and it therefore favours methods that reward epipolar consistency, such as the EAP and the gate. We report the number of epipolar-consistent matches and their fraction (precision), but treat pose accuracy as the primary metric.

Statistical testing.

Up to a negligible interpolation difference, the AUC at threshold T equals the mean over pairs of $\max(0, 1 - e/T)$, where e is the pose error. A difference in AUC between two methods is therefore a mean of paired per-pair differences. We test it with a two-sided sign-flip permutation test (20,000 flips) and summarize it with a percentile bootstrap [40] 95% confidence interval (CI; 5,000 resamples). Per-pair scores are averaged over the P RANSAC permutations. The pre-declared family consists of three hypotheses—EAP > SuperGlue (H1), PE compensation + EAP > SuperGlue (H2), and EAP > post-hoc gate (H3)—each tested on AUC@10° in two dataset groups (TUM-VI held-out, EuRoC) and four regimes, which gives 24 tests; the Holm procedure [41] controls the family-wise error rate. These tests treat pair evaluations as the unit of analysis, even though the same start frames recur across regimes and baselines and pairs within a sequence are correlated. To account for this dependence, we additionally report sequence-level analyses (per-sequence mean differences, an exact sign-flip test over sequences and a bootstrap over sequences). All other comparisons are descriptive.

4.4 Compared Methods

All methods use the same SuperPoint [3] detections and the same pretrained SuperGlue [1] (indoor weights, 20 Sinkhorn iterations, match threshold 0.2).

- **SuperGlue**: the unmodified matcher.
- **iso**, **maha**: pixel-domain rotation-only priors (Section 3.4; wide configuration, penalty-only).
- **PE comp. + maha**: PE compensation combined with the `maha` prior.
- **image warp**: image a is warped by the gyroscope rotation into the view of image b, SuperPoint and SuperGlue are run on the warped image, and the keypoints are mapped back.
- **post-hoc gate**: Section 3.7.
- **PE comp. only**: Section 3.6, no score prior.

- **EAP** (ours): Section 3.5.
- **PE comp. + EAP** (ours): both components.

In addition, we report the pose obtained from the gyroscope rotation and the known-rotation 2-point solver (**gRt**) for SuperGlue and for our matchers (Section 5.8). Finally, we evaluate LightGlue [2] (official implementation and SuperPoint weights, adaptive depth and pruning disabled) with and without the EAP on the same detections (Sections 3.8 and 5.6).

4.5 Deviations from the Protocol

The following deviations were recorded in the protocol log before the corresponding analyses were run, except where stated. (i) The selection criterion was the mean of the AUC@5°/10°/20° differences to SuperGlue over the development settings, under the constraint that no setting lose more than 2 points of AUC@10°. It was applied to 9 of the 12 planned development settings (full, low-budget and Gaussian-blur regimes): the motion-blur development runs did not finish in time, because OpenCV's RANSAC becomes slow at low inlier ratios. (ii) Having seen that the development gain grew with the baseline, we added a 1 s baseline as a secondary analysis before the held-out runs, with its own Holm family; the primary family was left unchanged. (iii) To limit run time, held-out pose estimates use $P = 5$ instead of 10 permutations. (iv) After the held-out runs, a code audit revealed that OpenCV's fisheye undistortion had mapped fisheye rays beyond 90° to their antipodes. This error affects the ray-space priors, PE compensation and image warping on TUM-VI, but not projective computations or pinhole cameras. We fixed it and re-ran all TUM-VI held-out experiments with the unchanged frozen configuration. The pooled TUM-VI gain of the EAP changed from +4.3 to +4.4 points (PE comp. + EAP: +3.0 to +3.2); the development selection itself had been made with the faulty mapping. (v) The control experiments of Section 5.5 were added after the held-out evaluation at the request of internal reviewers; they are descriptive. (vi) The settings for the qualitative figures (first held-out sequence of each dataset, motion blur, 600 ms) were fixed by rule after the summaries of two other held-out runs had been inspected as a sanity check, but before the runs used for the figures were available.

4.6 Implementation Details and Frozen Hyperparameters

The hyperparameters selected on `room1` (Section 5.2) and used unchanged on all held-out sequences are: EAP $\sigma_E = 0.5°$, $\kappa = 8$, offset $s = 8$, fallback rotation tolerance $\sigma_R = 6°$; PE comp. + EAP $\sigma_E = 1°$, $\kappa = 8$, $s = 8$. The previous pixel-domain priors use a wide configuration ($\sigma_\theta = 5°$, $\kappa = 2$ for `iso`; $\sigma_\theta = 0.6°$, $\sigma_t = 3$ ms, $\sigma_{\text{par}} = 6 + 150\,\Delta t$ px, $\kappa = 2$ for `maha`). The 2-point RANSAC uses 256 hypotheses and a Sampson threshold of 2 px/f inside the matcher and 1 px/f when used as a pose solver.

Run times are measured on an NVIDIA RTX 3090 with PyTorch 2.14 and OpenCV 5.0 (synchronized CUDA timing). Every reported number is generated by scripts from per-pair result files; each result file is accompanied by a metadata file that records all arguments, library versions and a hash of the source code.

5 Results

Tables numbered S1–S9 are provided in the Supplementary Material.

5.1 Calibration Chain and Gyroscope Accuracy

Before evaluating the priors, we checked the calibration chain. On the Hwangbo sequences, which lack ground truth, we measured for all 620 pairs how well the gyroscope transfers the RANSAC-consistent SuperGlue matches (Table S1 in the Supplementary Material). The rotation used by the method yields a median residual of 6.9 px, against 54.7 px without prediction and 108.3 px with the inverted rotation, and it gives the smallest residual in 99.5% of the pairs. On TUM-VI and EuRoC, we instead compared the integrated rotation with the ground truth (Table S2). After bias removal, the mean EuRoC gyroscope error per sequence and baseline stays at or below 0.40° even at 1 s; this figure may be optimistic, because the EuRoC ground truth is itself fused with the IMU. On TUM-VI, which offers no standstill window and where the bias is not removed, the error is at most 1.00°. Averaged over the held-out sequences, the error grows from 0.42° at 200 ms to 0.64° at 1 s on TUM-VI and from 0.05° to 0.18° on EuRoC; over the same baselines, the rotation between the frames grows from 11.9° to 43.1° (TUM-VI) and from 4.1° to 14.1° (EuRoC).

5.2 Development Sequence and Hyperparameter Selection

Table 2 shows the EAP grid on `room1`; like Table S3, it was computed before the fisheye fix of Section 4.5. Apart from a single exception, three trends hold across the grid. The gain increases as the angular tolerance σ_E decreases; it does not decrease with the truncation κ (the exception being $\sigma_E = 2°$ with $s = \kappa/2$); and in every cell it is larger with the reward-only offset ($s = \kappa$) than with the penalty-only offset ($s = 0$), consistent with the dustbin analysis of Section 3.3. The selected values of σ_E and κ lie at the boundary of the searched grid. Table S3 (Supplementary Material) compares the method families. On the development sequence, the pixel-domain rotation-only priors (`iso`, `maha`) reduce AUC, as do bearing-space rotation-only priors with a penalty-only offset; with a reward-only offset, a rotation-only prior is close to neutral (slightly negative). The post-hoc gate with the arc geometry is likewise neutral. By contrast, every EAP configuration in the grid improves the mean AUC@10°, and its two-sided (sign-agnostic) variant performs similarly to the one-sided arc. Applying the pre-declared criterion, we selected the EAP with $\sigma_E = 0.5°$, $\kappa = 8$, $s = \kappa$ and, for the combination with PE compensation, $\sigma_E = 1°$, $\kappa = 8$, $s = \kappa$, and froze these values for all held-out experiments.

Table 2 EAP hyperparameter grid on the development sequence (TUM-VI `room1`, 9 settings): mean ΔAUC@10° over SuperGlue in percentage points (pp) for angular tolerance σ_E, truncation κ and offset s. Blue bold/red italic: best/worst cell.

		mean ΔAUC@10 [pp]		
σ_E [°]	κ	$s = 0$	$s = \kappa/2$	$s = \kappa$
0.5	2	+0.89	+1.15	+1.57
	4	+1.01	+1.69	+1.78
	8	+1.10	+2.19	**+2.53**
1	2	+0.58	+0.83	+1.13
	4	+0.60	+1.16	+1.48
	8	+0.70	+1.50	+1.82
2	2	*+0.15*	+0.46	+0.57
	4	+0.18	+0.42	+0.95
	8	+0.18	+0.83	+1.44

Table 3 Held-out pose AUC@10° (%) per dataset and visual regime, pooled over the 200/400/600 ms baselines (mean over P=5 RANSAC point-order permutations). Best/worst per column in blue bold/red italic (baseline row not marked).

AUC [%]	**TUM-VI `room2`–`room6`**				**EuRoC (10 sequences)**			
Method	full	Gauss	128 kp	motion	full	Gauss	128 kp	motion
SuperGlue	66.3	62.3	55.8	19.7	43.8	43.8	36.8	16.0
iso (pixel)	*65.4*	*61.4*	*54.9*	19.7	*43.6*	43.6	*36.4*	15.7
maha (pixel)	65.9	61.8	*54.9*	19.8	43.9	*43.4*	36.6	*15.6*
PE comp. + maha	67.0	62.7	56.3	*19.1*	43.9	44.0	36.5	15.8
post-hoc gate	66.6	62.6	56.4	22.4	43.8	43.8	37.0	16.1
image warp	67.6	63.5	**58.9**	20.3	44.0	44.0	37.5	15.7
PE comp. only	67.5	63.5	56.8	20.9	44.3	44.3	36.5	15.8
EAP (ours)	**68.2**	**64.8**	**58.9**	**29.6**	**46.0**	**45.8**	**38.6**	**20.8**
PE comp. + EAP	68.0	64.5	58.8	25.6	45.1	44.8	37.5	17.3

5.3 Held-Out Evaluation

Table 3 reports the held-out AUC@10° obtained with the hyperparameters frozen on `room1`, pooled over the 200–600 ms baselines (n = 7,200 pair evaluations for TUM-VI and n = 14,400 for EuRoC), and Table 4 gives the correspondence statistics; Tables S8 and S9 (Supplementary Material) list AUC@5° and AUC@20° with their permutation standard deviations for TUM-VI and EuRoC. Counting ties at one decimal, the EAP attains the highest AUC@10° in 8 of the 8 dataset–regime combinations. Pooled over all regimes, it improves AUC@10° over SuperGlue by +4.4 points on TUM-VI (95% CI [+4.0, +4.7]) and by +2.7 points on EuRoC ([+2.5, +2.9]); AUC@5° and AUC@20° improve by +3.2/+4.9 (TUM-VI) and +1.6/+3.6 (EuRoC) points. The number of epipolar-consistent matches rises by +6.8% (TUM-VI) and +4.9% (EuRoC) and increases on 15 of the 15 sequences (Table S4). Their fraction, however, behaves differently on the two datasets. It increases on TUM-VI (70.4% → 72.1%) but decreases on EuRoC (79.5% → 78.0%; lower on 10 of the 10 EuRoC sequences), because on EuRoC the EAP adds more matches than it adds epipolar-consistent ones. The net effect on pose accuracy is nonetheless positive on both datasets. At the level of individual pairs, the EAP is not uniformly better: pooled over regimes, it raises the

Table 4 Correspondence statistics on held-out data (all regimes, 200–600 ms): mean number of matches and of epipolar-consistent matches per pair, and their fraction (precision, %), by the ground-truth epipolar test (Sampson distance < 3 px/f_x, no RANSAC). Blue bold/red italic: best/worst (SuperGlue and match counts not marked).

	TUM-VI			EuRoC		
Method	match	consist.	prec. [%]	match	consist.	prec. [%]
SuperGlue	173	128	70.4	239	196	79.5
iso (pixel)	167	*126*	71.3	236	*195*	**80.0**
maha (pixel)	170	127	70.9	237	*195*	79.7
PE comp. + maha	172	129	71.5	238	196	79.9
post-hoc gate	167	127	**72.5**	238	196	79.9
image warp	168	*126*	72.0	239	196	79.7
PE comp. only	175	130	71.0	239	196	79.6
EAP (ours)	183	**137**	72.1	257	**205**	78.0
PE comp. + EAP	186	135	*70.3*	260	202	*75.4*

per-pair score in 53% and lowers it in 32% of the TUM-VI pair evaluations (48% and 35% on EuRoC; the remainder are ties).

Table 5 summarizes the pre-declared hypothesis tests. After Holm correction at the pair level, all 24 of the 24 tests remain significant (largest adjusted $p = 0.008$). For H1 (EAP vs. SuperGlue), the gain ranges from +1.8 (EuRoC, low budget) to +9.9 points (TUM-VI, motion blur). For H3, the same arc geometry yields a larger improvement inside the assignment than as a post-hoc gate, which reuses the EAP's σ_E and κ and was not tuned separately. Pooled over regimes, the gate changes AUC@10° by +1.0 (TUM-VI) and +0.1 (EuRoC) points, against +4.4 and +2.7 for the EAP. H2 is significant as well, yet combining PE compensation with the EAP is weaker than the EAP alone in 8 of the 8 dataset–regime combinations (pooled: +3.2 vs. +4.4 on TUM-VI, +1.1 vs. +2.7 on EuRoC).

Pair evaluations are not independent, so Table 6 repeats the comparisons at the sequence level. The EAP improves AUC@10° on 15/15 sequences (exact sign-flip test over sequences, $p < 0.001$); a t-interval over the per-sequence means gives [+2.6, +6.1] (TUM-VI, 5 sequences) and [+1.9, +3.6] (EuRoC, 10 sequences). TUM-VI poses two limitations here. With only five sequences, a sequence-level sign test cannot reach $p < 0.0625$, and a bootstrap over five clusters is anti-conservative. The TUM-VI result therefore rests on the consistency across sequences (5/5) rather than on a small p-value. Per sequence, the gain ranges from +1.4 to +6.0 points, and its pair-level CI excludes zero in 15 of 15 sequences (Table S4). For nearly stationary pairs ($\|t\| < 1$ cm; 1.9% of TUM-VI and 7.8% of EuRoC pairs), the translation direction is ill-defined; excluding them changes the pooled gain to +4.5 and +2.9 points.

Table 5 Pre-declared hypothesis tests on held-out data (AUC@10°, pooled over 200/400/600 ms; H1: EAP vs. SuperGlue; H2: PE comp. + EAP vs. SuperGlue; H3: EAP vs. post-hoc gate). Δ: mean paired difference in percentage points (pp); CI: bootstrap 95% interval; W/L: pairs with higher/lower score; p: two-sided sign-flip permutation test; p_{Holm}: Holm-adjusted over the 24 tests.

Hyp.	Data	Regime	n	Δ [pp]	95% CI [pp]	W/L	p	p_{Holm}
H1	TUM	full	1800	+1.93	[+1.45, +2.44]	957/707	<0.001	0.001
H2	TUM	full	1800	+1.74	[+1.23, +2.27]	955/718	<0.001	0.001
H3	TUM	full	1800	+1.58	[+1.10, +2.06]	951/715	<0.001	0.001
H1	TUM	Gauss	1800	+2.56	[+2.01, +3.10]	983/668	<0.001	0.001
H2	TUM	Gauss	1800	+2.21	[+1.63, +2.77]	922/746	<0.001	0.001
H3	TUM	Gauss	1800	+2.19	[+1.68, +2.67]	968/685	<0.001	0.001
H1	TUM	128 kp	1800	+3.13	[+2.48, +3.79]	827/615	<0.001	0.001
H2	TUM	128 kp	1800	+3.03	[+2.34, +3.74]	896/679	<0.001	0.001
H3	TUM	128 kp	1800	+2.57	[+1.93, +3.19]	864/675	<0.001	0.001
H1	TUM	motion	1800	+9.88	[+9.02, +10.74]	1028/336	<0.001	0.001
H2	TUM	motion	1800	+5.90	[+5.16, +6.68]	856/450	<0.001	0.001
H3	TUM	motion	1800	+7.16	[+6.37, +7.97]	931/447	<0.001	0.001
H1	EuRoC	full	3600	+2.21	[+1.75, +2.68]	1838/1417	<0.001	0.001
H2	EuRoC	full	3600	+1.23	[+0.77, +1.70]	1698/1545	<0.001	0.001
H3	EuRoC	full	3600	+2.25	[+1.79, +2.72]	1851/1401	<0.001	0.001
H1	EuRoC	Gauss	3600	+2.04	[+1.58, +2.52]	1826/1421	<0.001	0.001
H2	EuRoC	Gauss	3600	+1.00	[+0.54, +1.45]	1702/1534	<0.001	0.001
H3	EuRoC	Gauss	3600	+1.98	[+1.53, +2.47]	1798/1446	<0.001	0.001
H1	EuRoC	128 kp	3600	+1.83	[+1.34, +2.32]	1674/1417	<0.001	0.001
H2	EuRoC	128 kp	3600	+0.70	[+0.18, +1.22]	1616/1485	0.008	0.008
H3	EuRoC	128 kp	3600	+1.64	[+1.13, +2.12]	1664/1438	<0.001	0.001
H1	EuRoC	motion	3600	+4.77	[+4.27, +5.28]	1621/832	<0.001	0.001
H2	EuRoC	motion	3600	+1.33	[+0.91, +1.78]	1231/1035	<0.001	0.001
H3	EuRoC	motion	3600	+4.71	[+4.22, +5.23]	1593/863	<0.001	0.001

Table 6 Sequence-level analysis of H1–H3 (AUC@10°, all regimes, 200–600 ms pooled). Pair CI: bootstrap over pair evaluations; Cluster CI: bootstrap over sequences; t-CI: t-interval over per-sequence means; Pos.: sequences with a positive mean difference; p_{seq}: exact sign-flip test over sequences.

Hyp.	Data	Δ [pp]	Pair CI [pp]	Cluster CI [pp]	t-CI [pp]	Pos.	p_{seq}
H1	TUM	+4.38	[+4.04, +4.71]	[+3.22, +5.33]	[+2.62, +6.13]	5/5	0.062
H1	EuRoC	+2.71	[+2.47, +2.96]	[+2.07, +3.45]	[+1.87, +3.56]	10/10	0.002
H1	All	+3.27	[+3.07, +3.47]	[+2.57, +3.99]	[+2.46, +4.07]	15/15	<0.001
H2	TUM	+3.22	[+2.89, +3.55]	[+2.03, +4.48]	[+1.23, +5.21]	5/5	0.062
H2	EuRoC	+1.06	[+0.83, +1.30]	[+0.66, +1.49]	[+0.56, +1.57]	10/10	0.002
H2	All	+1.78	[+1.59, +1.97]	[+1.12, +2.54]	[+0.97, +2.60]	15/15	<0.001
H3	TUM	+3.37	[+3.05, +3.69]	[+2.69, +4.05]	[+2.26, +4.49]	5/5	0.062
H3	EuRoC	+2.65	[+2.41, +2.89]	[+2.02, +3.38]	[+1.82, +3.47]	10/10	0.002
H3	All	+2.89	[+2.70, +3.08]	[+2.37, +3.45]	[+2.28, +3.50]	15/15	<0.001

5.4 Dependence on Rotation and Time Baseline

The benefit of the EAP grows with the inter-frame motion (Table 7, Fig. 3). On TUM-VI, the AUC@10° gain rises from +2.1 points at 200 ms to +6.5 at 600 ms and +9.9 at 1 s; on EuRoC, where rotations are smaller, it rises from +1.6 to +3.9 and +5.8. Binning all held-out pairs by the gyroscope rotation gives the same picture (Table S5), although rotation and dataset are partly confounded because EuRoC rotations are smaller: the gain is +1.8 points below 5°, +2.8 for 5°–15°, +4.9 for 15°–30° and +7.5 above 30°. The 1 s baseline was added as a secondary analysis, declared before the held-out runs. In this analysis (Table S6), all 16 of 16 tests are significant after

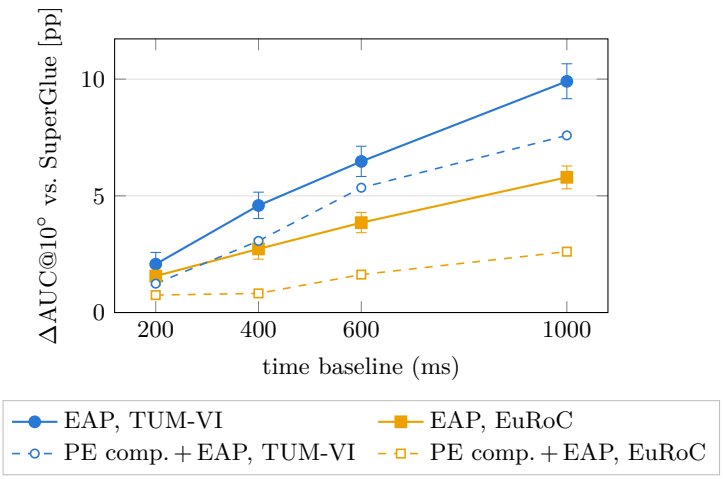


Fig. 3 AUC@10° gain over SuperGlue as a function of the time baseline (held-out, all four regimes pooled; error bars: bootstrap 95% CI).

Table 7 Effect of the time baseline (held-out, all four regimes pooled). Base.: time baseline; Rot.: mean gyroscope rotation; SG: SuperGlue AUC@10°; last two columns: ΔAUC@10° over SuperGlue with bootstrap 95% CI (PE comp. + EAP: Fig. 3).

Data	Base. [ms]	Rot. [°]	SG [%]	EAP [pp]	gate [pp]
TUM-VI	200	11.9	44.9	+2.1 [+1.6, +2.6]	+0.6 [+0.2, +0.9]
TUM-VI	400	22.5	53.7	+4.6 [+4.0, +5.2]	+1.0 [+0.7, +1.4]
TUM-VI	600	31.3	54.4	+6.5 [+5.8, +7.1]	+1.4 [+1.0, +1.8]
TUM-VI	1000	43.1	52.3	+9.9 [+9.2, +10.7]	+2.6 [+2.1, +3.1]
EuRoC	200	4.1	27.1	+1.6 [+1.2, +2.0]	+0.0 [-0.1, +0.1]
EuRoC	400	7.3	37.2	+2.7 [+2.3, +3.2]	+0.0 [-0.2, +0.2]
EuRoC	600	9.9	41.0	+3.9 [+3.4, +4.3]	+0.2 [+0.0, +0.4]
EuRoC	1000	14.1	42.7	+5.8 [+5.3, +6.3]	+0.3 [+0.1, +0.5]

Holm correction, with the largest gains under synthetic motion blur (+19.1 points on TUM-VI, +10.1 on EuRoC).

5.5 Control Experiments

We added the following experiments after the held-out evaluation to test alternative explanations of the gain (Section 4.5). They reuse the same held-out pairs and are descriptive (Table 8).

Is the gyroscope needed?

The gain could stem from the two-stage structure of the EAP itself: any relative-pose estimate obtained from the stage-1 matches already defines an epipolar arc. To test this, we estimated R and $\hat{t}$ from the stage-1 matches with the 5-point solver and applied the same prior. The vision-only variant inherits the σ_E and κ selected for the gyroscope-based EAP without re-tuning, so the comparison may overstate the contribution of the gyroscope. It improves AUC@10° by +1.5 (TUM-VI) and +0.5

Table 8 Post-hoc control experiments (held-out, all regimes, 200–600 ms): ΔAUC@10° over SuperGlue, relative change in the number of epipolar-consistent matches (Δcorr., %) and their fraction (prec., %). Blue bold/red italic: best/worst per column; the oracle (ground-truth pose) is not marked.

Variant	TUM-VI			EuRoC		
	ΔAUC@10 [pp]	Δcorr. [%]	prec. [%]	ΔAUC@10 [pp]	Δcorr. [%]	prec. [%]
SuperGlue	—	—	70.4	—	—	79.5
EAP (gyroscope R)	**+4.4**	**+6.8**	**72.1**	**+2.7**	**+4.9**	78.0
EAP without gyroscope	+1.5	+5.5	70.6	+0.5	+3.9	77.2
EAP, oracle R, t	+5.8	+9.7	74.0	+3.5	+6.2	79.3
rotation-only, σ_R=6°	+0.2	+1.7	67.0	-0.2	+1.5	73.6
rotation-only, σ_R=3°	*-0.5*	+0.7	67.5	*-0.6*	+1.2	74.1
constant offset +8	+0.8	+2.4	*66.1*	+0.1	+1.6	*73.3*
constant offset +2	+0.6	+1.3	69.2	+0.4	+0.9	77.8
match threshold 0.10	+0.3	*+0.5*	69.9	+0.2	*+0.4*	**78.7**
match threshold 0.05	+0.4	+0.7	69.5	+0.3	+0.6	78.3

points (EuRoC), against +4.4 and +2.7 for the EAP. Pooled over both datasets, the gyroscope-based EAP exceeds the vision-only variant by +2.5 points (cluster bootstrap CI [+1.9, +3.1]), on 15/15 sequences (sequence-level $p < 0.001$). Given the ground-truth relative pose, the same prior would reach +5.8 and +3.5 points; a better rotation or translation estimate would thus leave further room.

Is it only the match/no-match balance?

Because the reward-only offset raises candidates inside the arc band relative to the dustbin, it also increases the number of matches. Two geometry-free controls do the same: adding a constant to all scores, which is equivalent to lowering the dustbin score, and lowering the match threshold. Both increase the number of matches and decrease their epipolar consistency, yet change AUC@10° by at most +0.8 points on TUM-VI and +0.4 on EuRoC. The larger number of matches alone therefore does not explain the gain of the EAP.

Do rotation-only priors work with the EAP's offset and truncation?

Rotation-only priors in bearing space, given the same reward-only offset and $\kappa = 8$ as the EAP, change AUC@10° by between -0.6 and +0.2 points. The gap between these priors and the EAP is thus not explained by the offset or truncation, and is consistent with the EAP modelling parallax along the epipolar arc.

5.6 Transfer to LightGlue

To test whether the EAP transfers to a more recent matcher, we applied it to LightGlue (Section 3.8) on the same held-out pairs and SuperPoint detections, keeping the hyperparameters frozen at their SuperGlue values without re-tuning. We declared the tests before the LightGlue runs (H1-LG: EAP vs. LightGlue; H3-LG: EAP vs. post-hoc gate; 16 tests).

The effect is smaller than for SuperGlue (Table 9). Pooled over all regimes, the EAP raises the AUC@10° of LightGlue by +2.3 points on TUM-VI (CI [+2.0, +2.6])

and by +0.8 on EuRoC (CI [+0.7, +1.0]); the corresponding gains for SuperGlue are +4.4 and +2.7. After Holm correction, only 6 of the 16 pre-declared tests remain significant (Table 10). These are the tests under synthetic motion blur (+7.4 points on TUM-VI, +2.7 on EuRoC) and in the low-budget regime on TUM-VI (+1.3). In the remaining settings (unmodified images, Gaussian blur, and the low-budget regime on EuRoC), the differences lie between +0.1 and +0.3 points and are not significant. Per sequence, the EAP improves LightGlue on 14/15 sequences. As for SuperGlue, the gain depends on the gyroscope: the vision-only arc yields +0.3 and +0.1 points, and the gyroscope-based EAP outperforms it on 15/15 sequences.

Two properties of LightGlue are consistent with this smaller effect. First, its assignment has no dustbin column for the prior to shift, so the EAP cannot raise the number of matches in the same way; the number of epipolar-consistent matches grows by only +1.8% and +0.8%. Second, LightGlue is a stronger baseline on these data (AUC@10° 52.7 and 35.6 vs. 51.0 and 35.1 for SuperGlue). In practice, SuperGlue with the EAP (55.4 and 37.8) reaches a higher pooled AUC@10° than unmodified LightGlue on both datasets, without any training. Because we did not tune the prior for LightGlue, it remains open whether a different weighting of the prior relative to LightGlue's similarity scale would increase the effect.

Table 9 Transfer to LightGlue (held-out, 200/400/600 ms pooled, P=5): AUC@10° (%) per dataset and regime, and the pooled ΔAUC@10° over LightGlue. Best/worst per column in blue bold/red italic (baseline row not marked); "no gyro": arc estimated without the gyroscope.

AUC [%]	**TUM-VI**				**EuRoC**				**pooled Δ [pp]**	
Method	full	Gauss	128 kp	motion	full	Gauss	128 kp	motion	TUM	EuRoC
LightGlue	67.9	64.1	57.1	21.6	45.1	44.4	37.0	16.0	—	—
rotation-only	*67.7*	*64.0*	*57.1*	*21.0*	*44.8*	*44.4*	*36.8*	*16.0*	*-0.2*	*-0.1*
post-hoc gate	67.8	64.2	57.3	23.5	45.0	44.5	**37.1**	*16.0*	+0.5	+0.0
EAP, no gyro	**68.1**	64.3	57.3	22.1	45.0	*44.4*	36.9	16.4	+0.3	+0.1
EAP (ours)	**68.1**	**64.4**	**58.4**	**29.1**	**45.3**	**44.7**	**37.1**	**18.7**	**+2.3**	**+0.8**

Table 10 Pre-declared hypothesis tests for LightGlue on held-out data (AUC@10°, 200/400/600 ms pooled; H1-LG: EAP vs. LightGlue; H3-LG: EAP vs. post-hoc gate). Columns as in Table 5; Holm adjustment over these 16 tests.

Hyp.	**Data**	**Regime**	n	Δ [pp]	**95% CI [pp]**	**W/L**	p	p_{Holm}
H1	TUM	full	1800	+0.21	[-0.12, +0.54]	637/599	0.214	0.856
H3	TUM	full	1800	+0.31	[-0.04, +0.65]	651/619	0.076	0.605
H1	TUM	Gauss	1800	+0.27	[-0.08, +0.62]	571/582	0.137	0.820
H3	TUM	Gauss	1800	+0.20	[-0.15, +0.56]	590/604	0.296	0.888
H1	TUM	128 kp	1800	+1.26	[+0.88, +1.66]	282/191	<0.001	<0.001
H3	TUM	128 kp	1800	+1.09	[+0.69, +1.50]	391/290	<0.001	<0.001
H1	TUM	motion	1800	+7.44	[+6.67, +8.25]	893/378	<0.001	<0.001
H3	TUM	motion	1800	+5.53	[+4.77, +6.33]	844/447	<0.001	<0.001
H1	EuRoC	full	3600	+0.21	[-0.06, +0.49]	779/768	0.139	0.820
H3	EuRoC	full	3600	+0.30	[+0.04, +0.58]	804/751	0.029	0.294
H1	EuRoC	Gauss	3600	+0.30	[+0.02, +0.58]	840/744	0.041	0.367
H3	EuRoC	Gauss	3600	+0.25	[-0.02, +0.52]	838/760	0.079	0.605
H1	EuRoC	128 kp	3600	+0.11	[-0.11, +0.33]	397/380	0.354	0.888
H3	EuRoC	128 kp	3600	-0.00	[-0.26, +0.25]	506/489	0.980	0.980
H1	EuRoC	motion	3600	+2.70	[+2.31, +3.11]	1185/697	<0.001	<0.001
H3	EuRoC	motion	3600	+2.64	[+2.24, +3.04]	1182/712	<0.001	<0.001

5.7 Other Ways of Using the Gyroscope in the Matcher

The alternatives are also listed in Table 3. On held-out data, the pixel-domain rotation-only priors reduce AUC@10° slightly (`iso`: -0.7 on TUM-VI, -0.3 on EuRoC; `maha`: -0.4 and -0.2), consistent with the development results. PE compensation alone yields small gains on TUM-VI (+1.2) and almost none on EuRoC (+0.1), and combining it with the `maha` prior is close to neutral (+0.3 and -0.0). Image warping by the gyroscope rotation improves AUC@10° by +1.6 points on TUM-VI and +0.2 on EuRoC, at the cost of warping and re-detection. The post-hoc gate leaves the number of epipolar-consistent matches essentially unchanged (-0.4% and -0.0%) while slightly increasing their fraction. Of the interventions we evaluated, the EAP gives the largest pooled gain on both datasets. PE compensation combined with the EAP also gives significant, though smaller, gains, whereas the remaining interventions change pooled AUC@10° by at most +1.6 points.

5.8 Gyroscope in the Pose Solver

The gyroscope can also be exploited downstream: once the rotation is fixed to the gyroscope estimate, the 2-point solver recovers the translation direction. As Table 11 shows, the choice of solver affects pose AUC more than any of the matcher modifications we evaluated. With SuperGlue matches, AUC@10° rises from 51.0 to 69.0 on TUM-VI and from 35.1 to 60.3 on EuRoC (the EuRoC figures may be optimistic; see Section 4.1). Under this solver, the EAP changes AUC@10° by +0.5 points on TUM-VI (CI [+0.3, +0.7]) and +0.2 on EuRoC (CI [+0.0, +0.3]), and PE compensation + EAP by +0.8 and -0.1 points. Two conclusions follow. First, when the downstream module can use the gyroscope rotation directly, most of the pose benefit comes from the solver; the matcher-level gains in Tables 3–5 should therefore not be read as gains of a full VIO system. Second, the EAP helps most in pipelines whose pose solver does not use inertial data, where it closes about 24% (TUM-VI) and 11%

Table 11 Pose AUC@5°/10°/20° with the rotation-agnostic 5-point solver and with the gyroscope rotation plus the known-rotation 2-point solver, for SuperGlue and our matchers (held-out, 200–600 ms, all regimes). Blue bold/red italic: best/worst within each solver block (SuperGlue not marked).

AUC [%]		TUM-VI			EuRoC		
Solver	**Matcher**	@5	@10	@20	@5	@10	@20
5-point	SuperGlue	34.4	51.0	64.9	18.2	35.1	52.2
5-point	EAP	**37.6**	**55.4**	**69.8**	**19.8**	**37.8**	**55.8**
5-point	PE comp. + EAP	*37.0*	*54.2*	*68.2*	*18.9*	*36.1*	*53.7*
gyro R + 2-point	SuperGlue	56.5	69.0	79.3	42.5	60.3	73.3
gyro R + 2-point	EAP	*56.9*	*69.5*	*79.8*	**42.6**	**60.4**	**73.6**
gyro R + 2-point	PE comp. + EAP	**57.2**	**69.8**	**80.0**	*42.5*	*60.2*	*73.4*

(EuRoC) of the AUC@10° gap to the known-rotation solver. It also increases the number of epipolar-consistent matches; whether these additional matches help tracking or mapping was not evaluated.

5.9 Run Time

Table 12 lists run times measured in a single process after five warm-up pairs. Because the EAP does not run the GNN a second time, its GNN cost equals that of SuperGlue. The overhead lies instead in the assignment stage—a second Sinkhorn solve and the 2-point RANSAC—and amounts to 23 ms per pair in our implementation. This is comparable to the GNN pass itself (27 ms), because the RANSAC runs on the CPU in NumPy. The post-hoc gate incurs a similar, slightly smaller overhead, since it needs the same translation estimate. A GPU implementation of the 2-point RANSAC, which evaluates only a few hundred hypotheses, is expected to reduce this overhead; we did not optimize it. Image warping is the most expensive alternative, as it requires both warping and a second SuperPoint pass.

Table 12 Mean run time per pair (ms; single process, NVIDIA RTX 3090; TUM-VI `room2` and EuRoC `MH04`, 600 ms, 60 pairs each). GNN: GNN pass (for image warp including warping and re-detection); Assignment: all Sinkhorn solves plus, for EAP and gate, the 2-point RANSAC; blue bold/red italic: fastest/slowest (SuperGlue not marked).

Method	**GNN [ms]**	**Assignment [ms]**
SuperGlue	27.0	4.8
post-hoc gate	26.7	23.6
EAP (ours)	26.9	27.5
PE comp. only	24.0	**4.2**
PE comp. + EAP	**22.8**	*28.8*
image warp	*167.3*	4.3
iso (pixel)	27.3	5.7
maha (pixel)	27.6	5.6

5.10 Hwangbo Sequences (No Ground Truth)

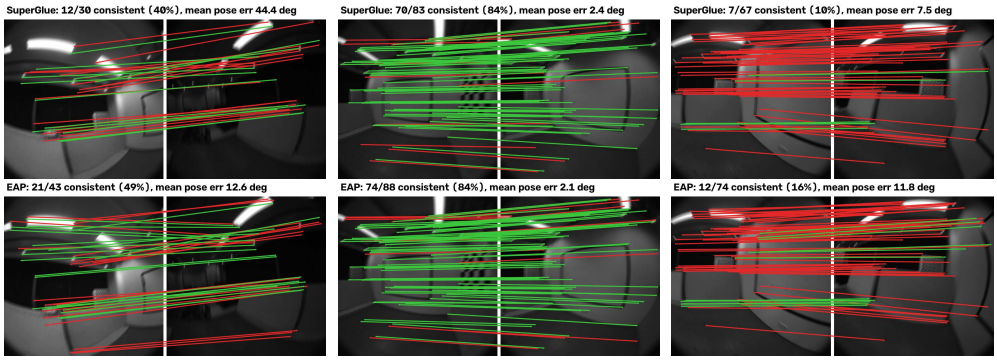
Without ground truth, the Hwangbo sequences allow us to measure only consistency (Table S7). The EAP increases the number of matches and, slightly, the absolute

number of matches consistent with the RANSAC essential matrix, whereas both the essential and the homography inlier ratios decrease. Both ratios are biased: the homography ratio favours rotation-consistent matches, and the essential-matrix ratio favours any epipolar-consistent set. With no ground truth available, these data therefore neither confirm nor refute the held-out findings. We report them for completeness, since this dataset is an established benchmark for gyroscope-aided tracking. In particular, the homography inlier ratio used previously favours rotation-only priors by construction and should not be used to compare such priors.

5.11 Qualitative Examples

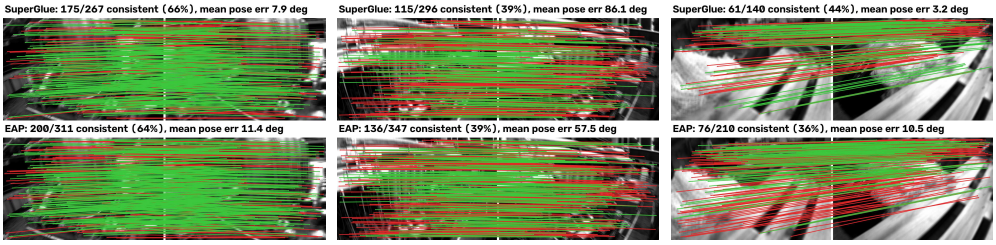
Figures 4 and 5 show the SuperGlue and EAP matches for the pairs chosen by the pre-declared rule: the pair with the largest rotation, a random pair, and the pair on which the EAP shows the largest *loss* relative to SuperGlue. Matches are coloured by the ground-truth epipolar test (green: epipolar-consistent, red: inconsistent), without any RANSAC step.

On TUM-VI `room2` (Fig. 4), SuperGlue and the EAP give (a) 12/30 vs. 21/43 epipolar-consistent matches and a mean pose error of 44.4° vs. 12.6°; (b) 70/83 vs. 74/88 epipolar-consistent matches and a mean pose error of 2.4° vs. 2.1°; (c) 7/67 vs. 12/74 epipolar-consistent matches and a mean pose error of 7.5° vs. 11.8°. On EuRoC `MH01` (Fig. 5), SuperGlue and the EAP give (a) 175/267 vs. 200/311 epipolar-consistent matches and a mean pose error of 7.9° vs. 11.4°; (b) 115/296 vs. 136/347 epipolar-consistent matches and a mean pose error of 86.1° vs. 57.5°; (c) 61/140 vs. 76/210 epipolar-consistent matches and a mean pose error of 3.2° vs. 10.5°. In both pairs with the largest loss, the EAP adds matches, among them epipolar-consistent ones, yet the pose estimate becomes worse; more epipolar-consistent matches therefore do not guarantee a better pose for an individual pair.



(a) largest rotation, 79.4° (b) random pair, 19.7° (c) largest EAP loss, 38.9°

Fig. 4 SuperGlue (top) and EAP (bottom) matches on TUM-VI `room2` under synthetic motion blur at 600 ms; green/red: consistent/inconsistent with the ground-truth epipolar geometry.



(a) largest rotation, 20.3° (b) random pair, 5.1° (c) largest EAP loss, 8.0°

Fig. 5 SuperGlue (top) and EAP (bottom) matches on EuRoC `MH01` under synthetic motion blur at 600 ms; green/red: consistent/inconsistent with the ground-truth epipolar geometry.

6 Discussion and Limitations

Why rotation-only priors do not help.

For distant structure the pure-rotation prediction is a good approximation. In the indoor and hall sequences used here, however, the parallax accumulated over 200–1000 ms amounts to several degrees (Fig. 2). To absorb it, a rotation-only prior needs a tolerance comparable to the typical distance between neighbouring keypoints, at which point the prior adds little information; a penalty-only offset additionally pushes matches into the dustbin. The epipolar arc, by contrast, removes the depth-dependent component of the residual, so the remaining tolerance is set mainly by the gyroscope and translation-direction errors. Our development grid (Table 2) is consistent with this explanation: the gain of the EAP increases as σ_E decreases, and the reward-only offset outperforms the penalty-only offset in every cell.

What the EAP depends on.

The EAP requires a translation direction, which it estimates from stage-1 matches. When the stage-1 matches are too few or the translation is unobservable (near-pure rotation), the method falls back to a loose rotation-only prior; when the estimate is wrong, the prior favours wrong correspondences. The arc construction further assumes a static scene, since independently moving objects violate the epipolar geometry of the camera motion. Such points are not removed, because the prior is truncated and reward-only; they only lose the advantage given to consistent points. Finally, the EAP was selected on a single development sequence, and both selected values ($\sigma_E = 0.5°$ and $\kappa = 8$) lie at the boundary of the searched grid. A tighter tolerance or a larger truncation may thus be preferable for well-calibrated sensors. The post-hoc baselines (gate, rotation-only priors), for their part, were tuned less extensively than the EAP.

Gyroscope quality.

The EuRoC gyroscope has a bias of about 0.08 rad/s around one axis. Without bias removal, the integrated rotation is off by 2.7°–3.5° after 600 ms (versus 0.04°–0.24° with it), several times larger than the selected σ_E. We removed the bias using a stand-still window; such a window is available in all EuRoC sequences but not in TUM-VI. On a real platform, the VIO back end typically estimates the bias continuously. Our

results therefore assume a bias-compensated gyroscope, and a poorly compensated one would require a larger σ_E.

Evaluation limitations.

Because the epipolar-consistency measure cannot detect wrong matches that lie on the correct epipolar line, it favours epipolar priors; for this reason we use pose accuracy as the primary metric. The EuRoC ground truth is fused with the IMU, so gyroscope errors and known-rotation results on EuRoC may be optimistic. Pair evaluations within a sequence are also correlated, which makes the sequence-level analysis (Table 6) the more conservative summary.

Synthetic degradations.

The blur regimes apply synthetic kernels to all images. Real motion blur, in contrast, is spatially varying and correlated with the rotation that the gyroscope measures, and we did not evaluate on real blurred sequences.

Scope.

We evaluated SuperGlue (indoor weights) and, without re-tuning, LightGlue with SuperPoint features. With LightGlue the effect was smaller and significant only under synthetic motion blur and, on TUM-VI, with a low keypoint budget. Its double softmax is invariant to the offset and its matchability is unaffected by the prior, so a matcher-specific weighting may be needed; we did not explore this option. In principle, the prior can also be added to the dual-softmax or OT scores of detector-free matchers such as LoFTR [12], which we did not test. Nor did we evaluate the matcher inside a complete VIO system or on consecutive frames at the full frame rate (50 ms). At that rate rotations are small and the expected gain is correspondingly smaller; the benefit of the matcher in a full system would depend on how the back end uses the correspondences.

7 Conclusions

We investigated how a gyroscope can be used inside a pretrained learned matcher without retraining. The most direct transfer of classical gyroscope-aided tracking, adding a prior around the pure-rotation prediction to SuperGlue's score matrix, did not improve relative-pose accuracy in our held-out evaluation. Our analysis attributes this to two factors: translation-induced parallax is large at the relevant baselines, and a penalty-only prior interacts with the dustbin. The EAP addresses both issues. It combines the gyroscope rotation with a translation direction estimated from the same GNN pass, constrains candidates across the epipolar arc while leaving them free along it, and only rewards consistent candidates. With hyperparameters frozen on a single development sequence, it improved pose AUC on all 15 held-out sequences of two datasets and increased the number of epipolar-consistent correspondences, and these gains grow with rotation. All pre-declared hypothesis tests remained significant after multiplicity correction at the pair level, and the improvement was consistent at the sequence level.

The evaluation also delimits the result. Using the gyroscope rotation in the pose solver yields a larger pose improvement than any of the evaluated matcher modifications, and once such a solver is in place, the EAP adds little to pose accuracy. Its benefit therefore lies mainly in pipelines whose pose solver does not use inertial data; whether its additional epipolar-consistent matches help tracking or mapping remains to be evaluated. On EuRoC, these additional matches come with a slightly lower fraction of epipolar-consistent matches. PE compensation and image warping gave smaller and less consistent gains. When transferred to LightGlue without re-tuning, the EAP gave smaller gains, significant only under synthetic motion blur and, on TUM-VI, with a low keypoint budget; even so, SuperGlue with the EAP reached a higher pose accuracy than unmodified LightGlue on both datasets. Several questions remain open: integration into a complete VIO back end with online bias estimation, evaluation on real (rather than synthetic) motion blur, the treatment of dynamic scenes, and the transfer of the EAP to other matchers with an explicit unmatched state.

Acknowledgements

Not applicable.

Declarations

Funding

This research received no external funding.

Conflict of interest

The author declares no conflict of interest.

Ethics approval and consent to participate

Not applicable. This study did not involve human participants or animals; it used the publicly available TUM-VI, EuRoC MAV, and Hwangbo/CMU IMU-KLT benchmark datasets.

Data availability

The datasets analyzed in this study are publicly available third-party datasets: the TUM-VI visual-inertial benchmark (Technical University of Munich) [38], the EuRoC MAV benchmark [39], and the Hwangbo/CMU IMU-KLT dataset [4, 5] (https://github.com/jomnipotent17/IMUTrack).

Code availability

The code, the frozen configuration, the evaluation protocol and all per-pair result files (with metadata recording arguments, library versions and source-code hashes) will be released publicly upon publication; until then they are available from the author upon request.

Author contribution

Conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, and visualization: U.A. The author has read and agreed to the published version of the manuscript.

References

[1] Sarlin, P.-E., DeTone, D., Malisiewicz, T., Rabinovich, A.: SuperGlue: Learning feature matching with graph neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4937–4946. IEEE, ??? (2020). https://doi.org/10.1109/CVPR42600.2020.00499 . Code: https://github.com/magicleap/SuperGluePretrainedNetwork

[2] Lindenberger, P., Sarlin, P.-E., Pollefeys, M.: LightGlue: Local feature matching at light speed. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 17581–17592. IEEE, ??? (2023). https://doi.org/10.1109/ICCV51070.2023.01616

[3] DeTone, D., Malisiewicz, T., Rabinovich, A.: SuperPoint: Self-supervised interest point detection and description. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW): Deep Learning for Visual SLAM Workshop, pp. 224–236. IEEE, ??? (2018). https://doi.org/10.1109/CVPRW.2018.00060

[4] Hwangbo, M., Kim, J.-S., Kanade, T.: Inertial-aided KLT feature tracking for a moving camera. In: 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1909–1916. IEEE, St. Louis, MO, USA (2009). https://doi.org/10.1109/IROS.2009.5354093 . PDF: https://www.cs.cmu.edu/~myung/IMU_KLT/KLT_iros_2009.pdf

[5] Hwangbo, M., Kim, J.-S., Kanade, T.: Gyro-aided feature tracking for a moving camera: Fusion, auto-calibration and GPU implementation. The International Journal of Robotics Research **30**(14), 1755–1774 (2011) https://doi.org/10.1177/0278364911416391 . PDF: https://www.cs.cmu.edu/~myung/publication/ijrr11.pdf

[6] Kneip, L., Chli, M., Siegwart, R.: Robust real-time visual odometry with a single camera and an IMU. In: Proceedings of the British Machine Vision Conference (BMVC), pp. 16–11611. BMVA Press, Dundee, UK (2011). https://doi.org/10.5244/C.25.16

[7] Troiani, C., Martinelli, A., Laugier, C., Scaramuzza, D.: 2-point-based outlier rejection for camera-IMU systems with applications to micro aerial vehicles. In: IEEE International Conference on Robotics and Automation (ICRA), pp. 5530–5536. IEEE, ??? (2014). https://doi.org/10.1109/ICRA.2014.6907672

[8] Li, H., Luo, K., Liu, S.: GyroFlow: Gyroscope-guided unsupervised optical flow learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 12849–12858. IEEE, ??? (2021). https://doi.org/10.1109/ICCV48922.2021.01263

[9] Li, H., Luo, K., Zeng, B., Liu, S.: GyroFlow+: Gyroscope-guided unsupervised deep homography and optical flow learning. International Journal of Computer Vision **132**(6), 2331–2349 (2024) https://doi.org/10.1007/s11263-023-01978-5 arXiv:2301.10018

[10] Cuturi, M.: Sinkhorn distances: Lightspeed computation of optimal transport. In: Advances in Neural Information Processing Systems (NeurIPS) 26, pp. 2292–2300. Curran Associates, Inc., ??? (2013). https://proceedings.neurips.cc/paper/2013/hash/af21d0c97db2e27e13572cbf59eb343d-Abstract.html

[11] Peyré, G., Cuturi, M.: Computational optimal transport: With applications to data science. Foundations and Trends in Machine Learning **11**(5-6), 355–607 (2019) https://doi.org/10.1561/2200000073

[12] Sun, J., Shen, Z., Wang, Y., Bao, H., Zhou, X.: LoFTR: Detector-free local feature matching with transformers. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 8918–8927. IEEE, ??? (2021). https://doi.org/10.1109/CVPR46437.2021.00881

[13] Chang, J., Yu, J., Zhang, T.: Structured epipolar matcher for local feature matching. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 6177–6186. IEEE, ??? (2023). https://doi.org/10.1109/CVPRW59228.2023.00657

[14] Kloepfer, D.A., Henriques, J.F., Campbell, D.: SCENES: Subpixel correspondence estimation with epipolar supervision. In: International Conference on 3D Vision (3DV), pp. 21–30. IEEE, ??? (2024). https://doi.org/10.1109/3DV62453.2024.00137

[15] Karpur, A., Perrotta, G., Martin-Brualla, R., Zhou, H., Araujo, A.: LFM-3D: Learnable feature matching across wide baselines using 3D signals. In: International Conference on 3D Vision (3DV), pp. 11–20. IEEE, ??? (2024). https://doi.org/10.1109/3DV62453.2024.00079

[16] Clerc, A., Quellmalz, M., Piening, M., Flotho, P., Kornhardt, G., Steidl, G.: HOT-POT: Optimal transport for sparse stereo matching. Journal of Mathematical Imaging and Vision **68**(4), 53 (2026) https://doi.org/10.1007/s10851-026-01336-3

[17] Bökman, G., Edstedt, J., Felsberg, M., Kahl, F.: Steerers: A framework for rotation equivariant keypoint descriptors. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4885–4895.

IEEE, ??? (2024). https://doi.org/10.1109/CVPR52733.2024.00467

[18] Lucas, B.D., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI), vol. 2, pp. 674–679. William Kaufmann, Vancouver, BC, Canada (1981). https://www.ijcai.org/Proceedings/81-2/Papers/017.pdf

[19] Baker, S., Matthews, I.: Lucas-Kanade 20 years on: A unifying framework. International Journal of Computer Vision **56**(3), 221–255 (2004) https://doi.org/10.1023/B:VISI.0000011205.11775.fd

[20] Li, P., Ren, H.: An efficient gyro-aided optical flow estimation in fast rotations with auto-calibration. IEEE Sensors Journal **18**(8), 3391–3399 (2018) https://doi.org/10.1109/JSEN.2018.2810060

[21] Huang, W., Liu, H.: A robust pixel-aware gyro-aided KLT feature tracker for large camera motions. IEEE Transactions on Instrumentation and Measurement **71**, 1–14 (2022) https://doi.org/10.1109/TIM.2021.3129493

[22] Mur-Artal, R., Montiel, J.M.M., Tardós, J.D.: ORB-SLAM: A versatile and accurate monocular SLAM system. IEEE Transactions on Robotics **31**(5), 1147–1163 (2015) https://doi.org/10.1109/TRO.2015.2463671

[23] 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) https://doi.org/10.1109/TRO.2016.2597321

[24] Qin, T., Li, P., Shen, S.: VINS-Mono: A robust and versatile monocular visual-inertial state estimator. IEEE Transactions on Robotics **34**(4), 1004–1020 (2018) https://doi.org/10.1109/TRO.2018.2853729 arXiv:1708.03852

[25] Leutenegger, S., Lynen, S., Bosse, M., Siegwart, R., Furgale, P.: Keyframe-based visual–inertial odometry using nonlinear optimization. The International Journal of Robotics Research **34**(3), 314–334 (2015) https://doi.org/10.1177/0278364914554813

[26] Campos, C., Elvira, R., Rodríguez, J.J.G., Montiel, J.M.M., Tardós, J.D.: ORB-SLAM3: An accurate open-source library for visual, visual–inertial, and multimap SLAM. IEEE Transactions on Robotics **37**(6), 1874–1890 (2021) https://doi.org/10.1109/TRO.2021.3075644

[27] Zhao, X., Liu, J., Wu, X., Chen, W., Guo, F., Li, Z.: Probabilistic spatial distribution prior based attentional keypoints matching network. IEEE Transactions on Circuits and Systems for Video Technology **32**(3), 1313–1327 (2022) https://doi.org/10.1109/TCSVT.2021.3068761

[28] Lu, Y., Zhao, Y., Wang, H., Ruan, Z., Liu, Y., Dong, Z., Yang, B.: Exploiting motion prior for accurate pose estimation of dashboard cameras. IEEE Robotics and Automation Letters **10**, 764–771 (2025) https://doi.org/10.1109/LRA.2024.3511381 arXiv:2409.18673 [cs.CV]

[29] Hamesse, C., Vlaminck, M., Luong, H., Haelterman, R.: Practical deep feature-based visual-inertial odometry. In: Proceedings of the 13th International Conference on Pattern Recognition Applications and Methods (ICPRAM), pp. 240–247. SCITEPRESS, ??? (2024). https://doi.org/10.5220/0012320200003654

[30] Fraundorfer, F., Tanskanen, P., Pollefeys, M.: A minimal case solution to the calibrated relative pose problem for the case of two known orientation angles. In: Computer Vision – ECCV 2010. Lecture Notes in Computer Science, vol. 6314, pp. 269–282. Springer, Berlin, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15561-1_20

[31] Fischler, M.A., Bolles, R.C.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM **24**(6), 381–395 (1981) https://doi.org/10.1145/358669.358692

[32] Torr, P.H.S., Zisserman, A.: MLESAC: A new robust estimator with application to estimating image geometry. Computer Vision and Image Understanding **78**(1), 138–156 (2000) https://doi.org/10.1006/cviu.1999.0832

[33] Baráth, D., Matas, J., Noskova, J.: MAGSAC: Marginalizing sample consensus. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 10189–10197. IEEE, ??? (2019). https://doi.org/10.1109/CVPR.2019.01044

[34] Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision, 2nd edn. Cambridge University Press, Cambridge, UK (2003). https://doi.org/10.1017/CBO9780511811685

[35] Nasypanyi, O., Rameau, F.: Pixel-Accurate Epipolar Guided Matching. arXiv preprint arXiv:2603.18401 (2026). https://doi.org/10.48550/arXiv.2603.18401 . https://arxiv.org/abs/2603.18401

[36] Kannala, J., Brandt, S.S.: A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses. IEEE Transactions on Pattern Analysis and Machine Intelligence **28**(8), 1335–1340 (2006) https://doi.org/10.1109/TPAMI.2006.153

[37] Nistér, D.: An efficient solution to the five-point relative pose problem. IEEE Transactions on Pattern Analysis and Machine Intelligence **26**(6), 756–770 (2004) https://doi.org/10.1109/TPAMI.2004.17

[38] Schubert, D., Goll, T., Demmel, N., Usenko, V., Stückler, J., Cremers, D.: The TUM VI benchmark for evaluating visual-inertial odometry. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1680–1687. IEEE, ??? (2018). https://doi.org/10.1109/IROS.2018.8593419

[39] 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) https://doi.org/10.1177/0278364915620033

[40] Efron, B., Tibshirani, R.J.: An Introduction to the Bootstrap. Monographs on Statistics and Applied Probability, vol. 57. Chapman and Hall/CRC, New York (1993). https://doi.org/10.1201/9780429246593

[41] Holm, S.: A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics **6**(2), 65–70 (1979)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementary.pdf