TRANSLATION AVERAGING AS GLOBAL DISCRETE INFERENCE ON THE SPHERE

Anonymous authors
Paper under double-blind review

ABSTRACT

Reconstructing a scene from images requires knowing where each photo was taken. Once camera orientations are fixed, the remaining step is *translation averaging*: estimating all camera positions given, for many camera pairs, the unit-length direction from one camera toward the other. These directions come from imperfect image matching, so a large fraction are not merely noisy but completely wrong (outlier). Existing solvers break down because none decides which measurements to trust: convex relaxations tolerate roughly a third of wrong directions, and robust local optimisers only refine their initial guess. Discrete optimisation is the classical remedy (give each variable a finite set of candidates and search globally for the combination explaining the most data, so the correct majority outvotes the wrong minority) but it needs a bounded candidate set, and camera positions live in unbounded 3D space. Our key idea is to change the question. A camera's position is unbounded, but the direction in which it is seen from a fixed reference point lies on the unit sphere, a bounded surface we cover with a fixed grid of cells. We plant K reference points and, for each, globally assign to every camera the cell in which it appears from that point. Each assignment pins every camera to a ray, and robustly intersecting a camera's K rays returns its 3D position. The proposed solver halves the median error of state-of-the-art solvers on synthetic graphs, and widens that margin to 4–$6\times$ at the outlier levels where convex relaxations break down. As a drop-in replacement for the global-positioning stage of the GLOMAP structure-from-motion pipeline, it improves or matches GLOMAP on every benchmark we test, wins end-to-end on ETH3D even after bundle adjustment, and, unlike its randomly initialised competitors, returns the same solution on every run. The code will be made public.

1 INTRODUCTION

Global structure-from-motion (Pan et al., 2024) recovers all camera poses at once over a view graph in two stages: first, rotation averaging fixes the absolute rotations (Govindu, 2004; Hartley et al., 2013). Then, *translation averaging* recovers the positions (Özyeşil et al., 2017). Given a unit direction $\tilde{\mathbf{t}}_{ij} \in S^2$ per edge, already in the world frame, the task is to estimate camera positions $\mathbf{c}_i \in \mathbb{R}^3$ with $\tilde{\mathbf{t}}_{ij} \propto \mathbf{c}_j - \mathbf{c}_i$ on as many edges as possible, up to global translation and scale. The rotation stage is comparatively settled (Pan et al., 2024). The translation stage, fed the noisy, outlier-laden directions of learned matchers, is the robustness bottleneck.

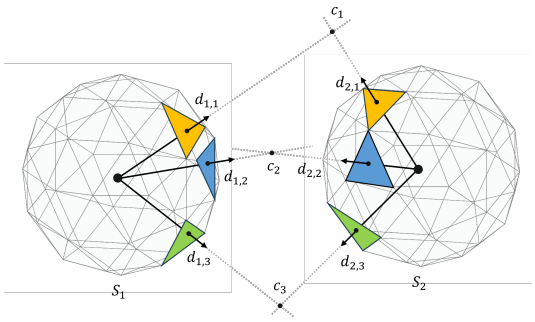


Figure 1: **Spherical direction labeling.** Each seed $\mathbf{S}_k \in \mathbb{R}^3$ sees every camera position $\mathbf{c}_i \in \mathbb{R}^3$ as a *direction* on a discretised sphere S^2. A Markov Random Field assigns the direction label $\mathbf{d}_{k,i}$ of every camera at once, in one global inference (triangles with the same color refer to the same viewpoint). Each label pins the camera to a ray from its seed (dotted); the rays of two seeds already intersect at $\mathbf{c}_i$; by using K seeds robust triangulation can remove mislabelled rays.

Two solver families dominate, and heavy outlier contamination defeats both. Convex and linear relaxations (Jiang et al., 2013; Moulon et al., 2013; Wilson & Snavely, 2014; Özyeşil & Singer, 2015; Cui et al., 2015; Cui & Tan, 2015; Goldstein et al., 2016; Sengupta et al., 2017)

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

carry guarantees but never decide which edges to trust: every wrong measurement keeps a vote, so its pull accumulates with contamination. Robust *local* optimisers, Huber/IRLS and the bilinear angle-based solver (BATA) of Zhuang et al. (2018) that GLOMAP runs, suppress large residuals but only descend from their initialisation, so a bad start is never escaped. Both degrade once outliers exceed roughly a third of the edges, and the diagnosis is the same on both sides: deciding which edges to trust is a *joint* question, and neither family ever searches that space of decisions.

The estimation toolbox has a classical cure for exactly this: global discrete MAP inference on a graphical model (Boykov et al., 2001; Kolmogorov, 2006), which reasons jointly over a label set rather than descending from one guess, so a consistent majority outvotes a contaminated minority. Crandall et al. (2011) brought this discrete-then-continuous template to structure-from-motion, but one obstacle has kept it out of translation averaging: discrete inference needs a *bounded* label space, and a camera position lives in unbounded $\mathbb{R}^3$. Voxelising space forces a scene-sized bounding box (Crandall et al., 2013), which must be guessed in advance.

So we change the question, and ask instead: "in which direction does camera i appear from seed k?" A camera's *position* is unbounded, but its *direction* seen from a fixed vantage point is a point on the unit sphere S^2, a compact manifold we discretise once and reuse for every camera and every vantage point. This poses the problem as a bounded, scene-independent labeling, and it still yields a metric position (Fig. 1): a label pins the camera to a ray, and K planted seeds give K rays whose outlier-robust intersection *is* the position. Each per-seed labeling is a sparse MAP problem on S^2 optimised globally by TRW-S. The K problems are decoupled and run in parallel. Cycling the seeds through several distances buys triangulation diversity for free, and a continuous polish (Huber, 1964; Pan et al., 2024) refines only within the basin the discrete stage has chosen. The construction comes with its analysis: the label pairs compatible with an edge form a one-dimensional locus on S^2, making the pairwise potential provably sparse (Prop. 1, Lemma 1); the unknown depths eliminate exactly (Cor. 1); and a quantisation bound (Prop. 2) says when the grid, not the outliers, limits accuracy, which is where the continuous refinement takes over.

Our contributions are (i) a **new formulation** of translation averaging as labeling each camera by its direction from planted seeds, the bounded, scene-independent label space that voxelising $\mathbb{R}^3$ cannot supply, which lets us search the inlier decisions globally instead of relaxing the objective or descending from a guess; and (ii) **global inference with an analysis**: decoupled per-seed MAP problems solved by TRW-S and fused by robust multi-view triangulation, with the locus, sparsity, depth-elimination and quantisation results that explain why it works. On synthetic graphs the solver is the strongest at almost every outlier rate, halving the median error of the state of the art and beating it by 4–6$\times$ under heavy contamination; as a drop-in positioner for GLOMAP it improves or matches the positioning stage on every benchmark we test and keeps its lead after bundle adjustment.

2 Related Work

Translation averaging and global SfM. Global SfM recovers all poses at once from two-view geometry (Govindu, 2001; Özyeşil et al., 2017), in contrast to incremental reconstruction (Snavely et al., 2006; Agarwal et al., 2009; Schönberger & Frahm, 2016), and underlies Theia (Sweeney et al., 2015) and GLOMAP (Pan et al., 2024). Its rotation stage is mature (Lie-algebraic (Govindu, 2004), robust (Chatterjee & Govindu, 2013; 2018) and certifiably optimal (Eriksson et al., 2018; Dellaert et al., 2020) solvers; see Hartley et al. (2013)) and supplies our inputs. Translation averaging follows two axes. Convex and linear relaxations (Wilson & Snavely, 2014; Özyeşil & Singer, 2015; Goldstein et al., 2016; Sengupta et al., 2017; Jiang et al., 2013; Cui & Tan, 2015; Cui et al., 2015; Moulon et al., 2013) are inexpensive and carry guarantees, but treat outliers statistically rather than combinatorially and degrade once they dominate. Robust local optimisation, Huber/IRLS and the bilinear angle-based solver of Zhuang et al. (2018) that GLOMAP adopts, with reweighting variants (Manam & Govindu, 2022), is only locally optimal, so the initialisation governs accuracy. Between the two, graduated non-convexity anneals a robust objective from a convex start (Sidhartha & Govindu, 2024), yet each stage remains a continuous descent, and triangle- or cycle-consistency methods (Shi & Lerman, 2018; Li et al., 2025; Fan et al., 2026) score or rescale edges by their agreement over 3-cycles before the solve; our triplet-closure confidences w_{ij} read the same signal, but only to weight edges, not to decide them. We depart from all of these: global *discrete* inference over a bounded label space makes an explicit combinatorial inlier decision, and continuous optimisation only polishes within the basin it selects.

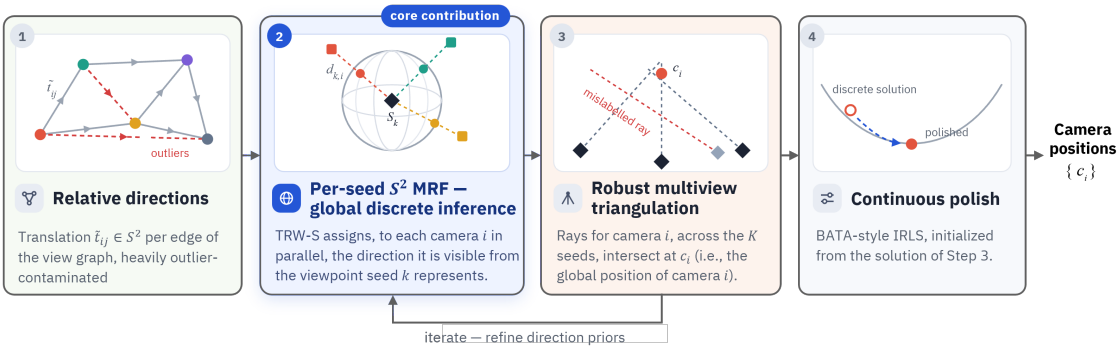


Figure 2: **Pipeline.** From noisy, outlier-contaminated relative directions $\tilde{\mathbf{t}}_{ij} \in S^2$ we plant K seeds (virtual viewpoints). For each seed k, a spherical Markov random field (MRF) solved globally by TRW-S (Kolmogorov, 2006) assigns every camera i the direction $\mathbf{d}_{k,i}$ in which it is seen from that seed. Robust multi-view triangulation fuses the K rays into positions $\mathbf{c}_i \in \mathbb{R}^3$, and a BATA-style IRLS polish (Zhuang et al., 2018) refines them.

Discrete MAP inference in geometric vision. Crandall et al. (2011; 2013) pioneered discrete inference for structure-from-motion: MAP labeling on an MRF minimised by loopy belief propagation (Felzenszwalb & Huttenlocher, 2006) with a continuous polish, the template we follow. Their translation MRF, however, discretises the unbounded position space directly, on a coarse grid bounded by known geotags – an input that we do not require. We keep the discretisation on a compact per-seed sphere, built once and reused, and recover the metric position only at the end by multi-seed triangulation. We solve each per-seed problem with the convergent, bound-providing TRW-S (Kolmogorov, 2006) rather than loopy BP (Szeliski et al., 2008), since our pairwise potential is not in general graph-cut-representable (Boykov et al., 2001).

Learned feed-forward reconstruction. DUSt3R (Wang et al., 2024), VGGT (Wang et al., 2025) and other recent methods regress pointmaps or poses directly. We see them as complementary directions: learned front-ends already supply our inputs, while the failure mode we address, a large fraction of grossly wrong edges over hundreds of views, is a global consistency problem for which explicit inference offers what a feed-forward pass does not: a consensus over the whole graph and an interpretable per-edge inlier decision.

3 Method

Rather than discretise a camera's unbounded position, we discretise the *direction* in which it appears from a planted *seed point*, a point on the compact sphere S^2 meshed once and reused. One direction label pins a camera to a ray; several seeds pin it to where their rays meet. Recovering one seed's labels is a discrete maximum-a-posteriori (MAP) problem on S^2 optimised globally by tree-reweighted message passing, and this global discrete inference is what buys robustness to gross direction outliers. We develop the model along the pipeline of Fig. 2; proofs are in Appendix B, and Algorithm 1 (Appendix A) provides the pseudo-code of the procedure.

Problem statement and gauge fixing. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a view graph with $N{=}|\mathcal{V}|$ cameras. Each undirected edge $(i, j) \in \mathcal{E}$ carries a unit relative-direction measurement $\tilde{\mathbf{t}}_{ij} \in S^2$, already rotated into the world frame by the separately estimated absolute rotations. We seek positions $\mathbf{c}_i \in \mathbb{R}^3$ with

$$\tilde{\mathbf{t}}_{ij} \;\propto\; \mathbf{c}_j - \mathbf{c}_i, \qquad \mathbf{c}_j - \mathbf{c}_i \neq \mathbf{0}, \tag{1}$$

on as many edges as possible. Directions are blind to a global translation and scale, $\mathbf{c}_i \mapsto s\,\mathbf{c}_i + \mathbf{t}_0$, so we fix these four degrees of freedom up front by choosing a high-confidence *seed edge* $(i^\star, j^\star)$ (ranked by triplet-closure confidence) and pinning its endpoints,

$$\mathbf{c}_{i^\star} = \mathbf{0}, \qquad \mathbf{c}_{j^\star} = \tilde{\mathbf{t}}_{i^\star j^\star}, \tag{2}$$

which also takes the seed edge's measured direction as exact.

3.1 From positions to directions on the sphere

Discretisation needs a bounded label space, and a raw position offers none without a bounding box for the scene. We obtain one by changing the question from "where is camera i?" to "in which direction does camera i appear from a given point?". We plant K *seed points* $\{\mathbf{S}_k\}_{k=1}^{K} \subset \mathbb{R}^3$ in the gauged frame and give each camera i, for each seed k, a *direction label*

$$\mathbf{d}_{k,i} := \frac{\mathbf{c}_i - \mathbf{S}_k}{\|\mathbf{c}_i - \mathbf{S}_k\|} \in S^2. \tag{3}$$

Every label lives on the same compact 2-manifold however far the camera is, so a single discretisation suffices: we subdivide an icosahedron ℓ times into V cells $\{\mathbf{v}_a\}_{a=1}^{V}$; at level ℓ=3, V=642 cells of angular (covering) radius $\rho \approx 4°$. A label fixes the position up to an unknown depth $r_{k,i} > 0$ along the ray,

$$\mathbf{c}_i = \mathbf{S}_k + r_{k,i}\,\mathbf{d}_{k,i}, \tag{4}$$

so one seed cannot place a camera but two or more can: the camera is where the rays meet. Each seed thus yields an independently solvable labeling, and the seeds are reconciled into positions only at the end. What remains is to say, for one seed, which labelings of the whole graph are consistent with the measurements.

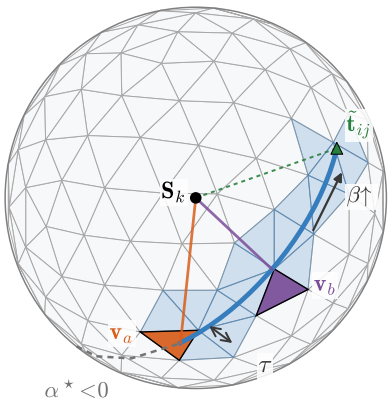


Figure 3: **Pairwise compatibility on S^2.** Cells are candidate directions seen from seed $\mathbf{S}_k$. With camera i labelled $\mathbf{v}_a$, the compatible labels $\mathbf{v}_b$ form the geodesic arc swept as the baseline β grows, ending at $\tilde{\mathbf{t}}_{ij}$ (Prop. 1). Only cells within τ of that arc (shaded) are connected by ψ_{ij}, hence its sparsity (App. B.3); the backward branch would need $\alpha^\star$<0.

3.2 Pairwise compatibility on S^2

Within one seed's labeling, two cameras interact only through the edge that connects them, scored by the pairwise potential $\psi_{ij}(\mathbf{v}_a, \mathbf{v}_b)$ over the candidate labels $\mathbf{v}_a, \mathbf{v}_b$ of cameras i and j. The intuition is geometric (Fig. 3): camera i lies *somewhere* along ray $\mathbf{v}_a$ and camera j *somewhere* along ray $\mathbf{v}_b$; the labels are *compatible* if positive depths exist that place j in the measured direction $\tilde{\mathbf{t}}_{ij}$ from i. Only the existence of such depths matters, not their values. Writing $\mathbf{c}_i = \mathbf{S}_k + r_{k,i}\mathbf{v}_a$ and $\mathbf{c}_j = \mathbf{S}_k + r_{k,j}\mathbf{v}_b$, the seed cancels in $\mathbf{c}_j - \mathbf{c}_i$ and (1) becomes a relation between the rays alone,

$$r_{k,j}\,\mathbf{v}_b - r_{k,i}\,\mathbf{v}_a \;=\; \beta\,\tilde{\mathbf{t}}_{ij}, \quad \beta = \|\mathbf{c}_j - \mathbf{c}_i\| > 0, \tag{5}$$

with β the unknown baseline length. The labels satisfying (5) for *some* positive depths admit an exact, one-dimensional description, the fact on which the whole construction rests.

Proposition 1 (Compatibility locus). *Fix $\tilde{\mathbf{t}}_{ij} \in S^2$ and a candidate label $\mathbf{v}_a \in S^2$ with $\mathbf{v}_a \neq \pm\tilde{\mathbf{t}}_{ij}$. There exist depths $r_{k,i}, r_{k,j} > 0$ and a baseline $\beta > 0$ satisfying* (5) *if and only if $\mathbf{v}_b$ lies on the open geodesic arc*

$$\mathcal{A}(\mathbf{v}_a, \tilde{\mathbf{t}}_{ij}) \;=\; \left\{ \tfrac{\mathbf{v}_a + s\,\tilde{\mathbf{t}}_{ij}}{\|\mathbf{v}_a + s\,\tilde{\mathbf{t}}_{ij}\|} \;:\; s \in (0,\infty) \right\}, \tag{6}$$

the minor arc of the great circle through $\mathbf{v}_a$ and $\tilde{\mathbf{t}}_{ij}$, traversed from $\mathbf{v}_a$ (excluded, $s\to0$) toward $\tilde{\mathbf{t}}_{ij}$ (excluded, $s\to\infty$).

What the locus means. Stand at the seed and suppose camera i lies in direction $\mathbf{v}_a$; its distance and the edge length β are unknown, but the edge's direction $\tilde{\mathbf{t}}_{ij}$ is not. A very short edge puts j almost on top of i, so $\mathbf{v}_b \to \mathbf{v}_a$; a very long one puts j far out along the edge, where i's exact place stops mattering, so $\mathbf{v}_b \to \tilde{\mathbf{t}}_{ij}$. In between, $\mathbf{v}_b$ is the normalisation of $\mathbf{v}_a + s\,\tilde{\mathbf{t}}_{ij}$, which never leaves the plane of those two vectors, and a plane through the origin cuts the sphere in a great circle. Two consequences matter: the compatible set is a *curve* in a two-dimensional label space, which is what makes the model sparse; and it stops at $\tilde{\mathbf{t}}_{ij}$, since going further would put j *behind* i, a different measurement. That is the forward branch; the rest of the circle is excluded. The proof is in Appendix B.1.

Eliminating the unknown depths. With $(\mathbf{v}_a, \mathbf{v}_b)$ fixed by the inference, the only unknowns in (5) are the depths. Scale is gauged, so set $r_{k,i}$=1 and let $\alpha := r_{k,j}$ be the *depth ratio*; compatibility

asks whether some $\alpha > 0$ makes $\alpha \mathbf{v}_b - \mathbf{v}_a$ positively parallel to $\tilde{\mathbf{t}}_{ij}$. Parallelism is a vanishing cross product, *linear* in α,

$$\alpha\, \mathbf{q}_b = \mathbf{q}_a, \qquad \mathbf{q}_a := \tilde{\mathbf{t}}_{ij} \times \mathbf{v}_a, \;\; \mathbf{q}_b := \tilde{\mathbf{t}}_{ij} \times \mathbf{v}_b, \tag{7}$$

whose least-squares solution is the scalar projection

$$\alpha^\star = \mathbf{q}_b \cdot \mathbf{q}_a / \|\mathbf{q}_b\|^2. \tag{8}$$

The elimination is exact: no compatible pair is ever misscored (for $\mathbf{q}_b \neq \mathbf{0}$; Cor. 1).

Corollary 1 (Exact depth elimination)**.** *Assume $\mathbf{q}_b \neq \mathbf{0}$. Then $\alpha^\star$ in* (8) *is the unique minimiser of $\|\tilde{\mathbf{t}}_{ij} \times (\alpha \mathbf{v}_b - \mathbf{v}_a)\|^2$ over $\alpha \in \mathbb{R}$, and $(\mathbf{v}_a, \mathbf{v}_b)$ admits exact positive depths in the sense of Prop. 1 if and only if $\alpha^\star > 0$ and $\theta := \angle(\alpha^\star \mathbf{v}_b - \mathbf{v}_a,\, \tilde{\mathbf{t}}_{ij}) = 0$.*

Scoring the residual. For a noisy pair the residual angle θ is nonzero and is exactly the quantity to score: how far the displacement $\mathbf{p}_{ij} := \alpha^\star \mathbf{v}_b - \mathbf{v}_a$ implied by the labels deviates from the measured direction. The potential rewards small θ with a Gaussian, hard-gated by an inlier threshold and the forward branch,

$$\psi_{ij}(\mathbf{v}_a, \mathbf{v}_b) \;=\; \exp\!\Big(-\frac{\theta^2}{2\sigma^2}\Big)\, \mathbb{1}[\alpha^\star > 0]\, \mathbb{1}[\theta < \tau], \tag{9}$$

with τ the angular inlier threshold and $\sigma = \tau/5$. The gate $\theta < \tau$ rejects pairs off the tube, including the backward branch where $\theta \approx \pi$; $\alpha^\star > 0$ rejects pairs needing a negative depth; coincident cells are excluded as degenerate.

Sparsity. Because the exactly compatible set is a curve, we build the gated potential only on a thin tube around it (the gate alone is looser near $\mathbf{v}_a \approx -\tilde{\mathbf{t}}_{ij}$; App. B.3). Let $\bar{k}$ be the number of cells $\mathbf{v}_b$ with nonzero $\psi_{ij}(\mathbf{v}_a, \mathbf{v}_b)$ for a fixed $\mathbf{v}_a$, the nonzeros per row of the $V \times V$ compatibility table, which we build sparsely by walking the curve and ball-querying the grid around each sample. At level 3 (V=642) we measure $17 \leq \bar{k} \leq 67$, growing with the arc length $\angle(\mathbf{v}_a, \tilde{\mathbf{t}}_{ij})$, so at most $\sim$10% of the cells are active (Fig. 3 draws a coarser grid for legibility); Appendix B.3 bounds $\bar{k}$ and matches these counts.

Seed placement and scale diversity. The first two seeds coincide with the gauge cameras ($\mathbf{S}_1$=$\mathbf{0}$, $\mathbf{S}_2$=$\tilde{\mathbf{t}}_{i^\star j^\star}$) and come with exact hard locks. The remaining $K-2$ are *virtual* seeds on canonical unit directions $\{\hat{\mathbf{u}}_m\}_{m=1}^{K-2}$ in the plane perpendicular to the seed edge (the axes $\pm\mathbf{e}_1, \pm\mathbf{e}_2$ of an orthonormal basis of $\tilde{\mathbf{t}}_{i^\star j^\star}^{\perp}$ and two diagonals), chosen for ray diversity; the anti-pole $-\tilde{\mathbf{t}}_{i^\star j^\star}$ is skipped as collinear with both gauge cameras. How *far* to place them is genuinely uncertain: the gauge fixes the seed edge at unit length, but the camera cluster may span much more or less. Too close and seeds fall inside the cluster where rays fan out; too far and rays become near-parallel; both degrade the triangulation conditioning (App. B.4). Rather than estimate the right scale we *cover* it with a small set of distances $\mathcal{A}_s = \{a_1, \ldots, a_L\}$ that the virtual seeds cycle through,

$$\mathbf{S}_{m+2} = a_{((m-1) \bmod L)+1}\, \hat{\mathbf{u}}_m, \quad m = 1, \ldots, K-2. \tag{10}$$

With L=3 and $\mathcal{A}_s = \{0.5, 1.0, 2.0\}$ the virtual seeds span a $4\times$ range, so the triangulation blends short, medium and long baselines at no extra cost: each seed sits at one distance, so we solve K MRFs, not $K \times L$.

3.3 Anchoring the labeling

The pairwise term constrains only *relative* geometry, so by itself it cannot single out one labeling; the residual freedom is the gauge group of the original problem, resurfacing for each seed.

Remark 1 (Gauge freedom of the per-seed model). (i) By (4), one seed's labeling fixes each camera only up to its depth along the labeled ray. (ii) The data term is invariant under $\mathbf{c}_i \mapsto s\,\mathbf{c}_i + \mathbf{t}_0$ ($s > 0$), yet the transformed configuration induces a *different* direction field from the same seed. Without anchors, the per-seed objective has a 4-parameter family of equally consistent labelings, which appear on the grid as near-ties.

Two ingredients pin this down. **Hard locks:** labels that follow from the gauge (2) are fixed exactly and removed from message passing by binary masks: $\mathbf{S}_1$ coincides with $\mathbf{c}_{i^\star}$, locking $\mathbf{d}_{1,j^\star} = +\tilde{\mathbf{t}}_{i^\star j^\star}$; $\mathbf{S}_2$ coincides with $\mathbf{c}_{j^\star}$, locking $\mathbf{d}_{2,i^\star} = -\tilde{\mathbf{t}}_{i^\star j^\star}$; and every virtual seed $\mathbf{S}_k$ (k≥3) sees both gauge cameras in exact directions $\mathbf{d}_{k,i^\star} = -\mathbf{S}_k/\|\mathbf{S}_k\|$ and $\mathbf{d}_{k,j^\star} = (\tilde{\mathbf{t}}_{i^\star j^\star} - \mathbf{S}_k)/\|\cdot\|$.

Pinning the two gauge cameras removes exactly the four degrees of freedom of Remark 1(ii); a camera that *coincides* with its seed has no direction and is dropped from that seed's MRF. **Soft direction priors:** the locks say nothing about interior cameras, so for each remaining camera i we take a coarse estimate $\hat{\mathbf{c}}_i$ from a cheap continuous pre-solve (the BATA-style IRLS of Zhuang et al. (2018) that GLOMAP runs, applied once to the raw edges), project it to a direction from each seed, and add a broad Gaussian unary that gently favours the corresponding cell,

$$\phi_{k,i}(\mathbf{v}) = w_u \exp\Big(-\tfrac{1}{2\sigma_u^2}\,\angle^2\big(\mathbf{v},\ \tfrac{\hat{\mathbf{c}}_i-\mathbf{S}_k}{\|\hat{\mathbf{c}}_i-\mathbf{S}_k\|}\big)\Big), \tag{11}$$

with width $\sigma_u = \tau + 1.5\rho$ and peak height $w_u = 0.1$: broad enough to break the near-ties of Remark 1 without overriding the data-driven pairwise term. What matters here is the weight, not the estimate. The prior only has to select among the labelings the pairwise term already deems equally good, so replacing the pre-solve by a breadth-first propagation over high-confidence edges leaves accuracy similar, while letting the prior speak louder degrades it sharply under heavy contamination, the pairwise consensus being what rejects outliers (App. C).

3.4 Per-seed inference and iterative refinement

For seed k, we seek the labeling maximising the unary and pairwise evidence,

$$\arg\max_{\{\mathbf{v}_{k,i}\}} \sum_{i\in\mathcal{V}} \phi_{k,i}(\mathbf{v}_{k,i}) + \sum_{(i,j)\in\mathcal{E}} w_{ij}\,\psi_{ij}(\mathbf{v}_{k,i},\mathbf{v}_{k,j}), \tag{12}$$

where labels range over the grid cells and w_{ij} is the edge's triplet-closure confidence (App. C). Both terms are bounded rewards: a label pair outside the tube of (9), or on the backward branch, earns nothing for that edge, so an outlier edge can lower a labeling's score by at most w_{ij} and never veto it. This is what lets a compatible majority outvote a contaminated minority. The hard locks of Sec. 3.3 are the only infinite penalties. We solve (12) with tree-reweighted message passing (TRW-S; Kolmogorov, 2006). The K MAP problems share only the grid and the measurements, so they are fully *decoupled* and run in parallel; decoding seed k assigns each camera a direction $\hat{\mathbf{d}}_{k,i}$.

Refinement. The grid quantises each direction to a cell of radius ρ, which costs precision when the camera cluster spans many gauge units, so we refine the per-seed direction field by a continuous IRLS descent initialised at the TRW-S solution.

3.5 Fusing the seeds: robust multi-view triangulation

Each seed supplies one ray per camera, $(\mathbf{S}_k, \hat{\mathbf{d}}_{k,i})$, and the camera sits where the rays meet. Since an MRF can mislabel a direction, some rays are wrong, so we recover the position with the robust multi-view triangulation of COLMAP (Schönberger & Frahm, 2016): every pair of rays proposes a candidate, the largest set of rays agreeing with a candidate to within $\tau + \rho$ is kept, and the position is refit on that inlier set $\mathcal{I}$ by point-to-ray least squares (App. C).

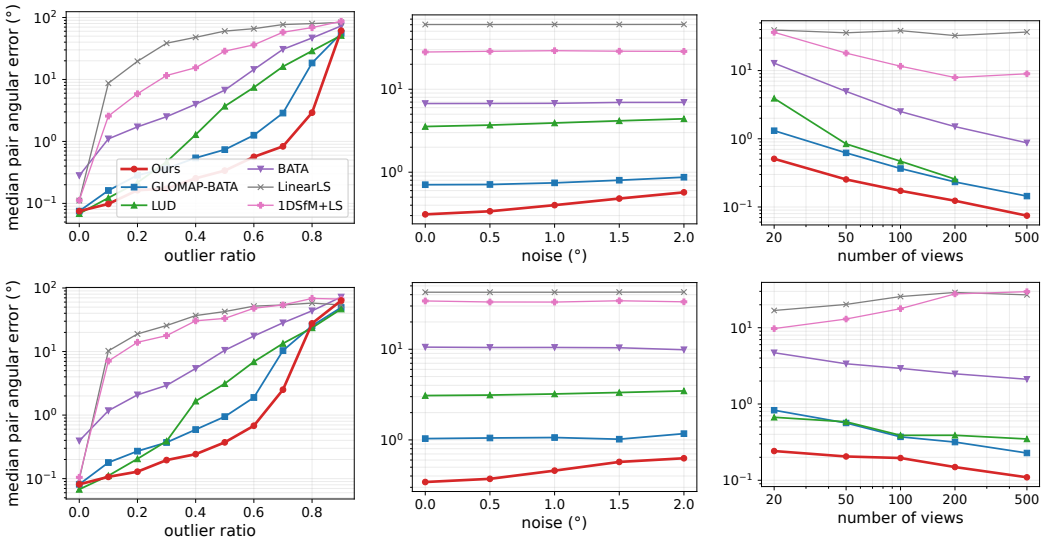
4 Experiments

4.1 Outlier robustness

Setup. Ground-truth positions are drawn uniformly from $[-1, 1]^3$ under two topologies: *dense* Erdős–Rényi graphs (p=0.5) and *sparse* k-nearest-neighbour graphs (k=20), the latter closer to SfM collections. Each clean direction is perturbed by von-Mises–Fisher noise of concentration $\kappa = 1/\sigma^2$ (σ the angular noise scale in degrees) and replaced by a uniformly random unit vector with probability ρ_{out}, the *outlier rate*. Unless stated otherwise, N=100 views, σ=0.5° and ρ_{out}=0.3. We report the median pair angular error $\angle\big(\mathbf{c}_j^{\text{est}} - \mathbf{c}_i^{\text{est}},\ \mathbf{c}_j^{\text{gt}} - \mathbf{c}_i^{\text{gt}}\big)$ averaged over 20 random seeds.

Baselines. GLOMAP-BATA (Huber + Ceres on the BATA direction error) is the production positioner of Pan et al. (2024), run from a random initialisation. LUD (Özyeşil & Singer, 2015) is the least-unsquared-deviations convex relaxation; BATA and BATA-IRLS are the bilinear angular solver (Zhuang et al., 2018) without and with Geman-McClure reweighting; LinearLS is the non-robust cross-product least-squares solver, and 1DSfM+LS runs it on the edges surviving the filter of Wilson & Snavely (2014). Ours uses K=8 seeds.

Results. Figure 4a sweeps the outlier rate from 0 to 90%. Ours is the strongest on *both* topologies at every rate up to a phase transition near 80–90% on dense graphs and 70–80% on sparse ones. On



(a) vs. outlier ratio (noise 0.5°)

(b) vs. noise (50% outliers)

(c) vs. views (30% outliers, 0.5°)

Figure 4: **Synthetic robustness.** Median pair angular error on synthetic graphs. *Top row:* dense (ER-p=0.5); *bottom row:* sparse (kNN-k=20). Each column sweeps one axis with the others fixed, as noted; N=100 in (a) and (b). The legend in the top-left panel applies throughout.

dense graphs it reaches 0.16° at ρ_{out}=0.2, 0.34° at 0.5 and 0.83° at 0.7, against 0.28/0.74/2.87° for GLOMAP-BATA, a 2–3.5× margin that widens with contamination, and an order of magnitude better than LUD beyond ρ_{out}=0.4. On sparse graphs the ordering is the same (0.24/0.37/0.68° at ρ_{out}=0.4/0.5/0.6 vs. 0.59/0.94/1.89° for GLOMAP-BATA and 1.66/3.12/6.90° for LUD) but the transition comes earlier: from ρ_{out}=0.8 too few inlier edges per view remain for any consensus and the gracefully degrading baselines overtake us (27.7° vs. 25.4° for GLOMAP-BATA and 23.6° for LUD). Figure 4b fixes the outlier rate at 50% and sweeps the noise: both curves are nearly flat, ours rising from 0.31° to 0.57° on dense and from 0.34° to 0.63° on sparse graphs across $\sigma \in [0, 2]°$, a steady 1.5–3× lead over GLOMAP-BATA (0.71 → 0.87° dense, 1.03 → 1.17° sparse) and an order of magnitude over LUD. Figure 4c varies the problem size (30% outliers, 0.5°): the dense-graph margin over GLOMAP-BATA stays at ∼ 2× across $N \in \{20, \ldots, 500\}$ while our error falls monotonically to 0.08° at N=500, so the advantage is not tied to one problem size.

Complexity and runtime. Per seed, the compatibility construction and a TRW-S sweep each cost $O(|\mathcal{E}|\, V\, \bar{k})$, with $\bar{k}$ the number of cells a label interacts with (App. B.3), against $O(|\mathcal{E}|\, V^2)$ for a dense interaction. With V=642, $\bar{k} \approx 30$ and K=8 MRFs in parallel, the reference implementation costs 10 s at N=20, 307 s at N=100 and 1103 s at dense N=500. Sparse graphs are 2–4× cheaper at equal N since the compatibility tensor scales with the edge count. The continuous baselines run in seconds: our accuracy is bought with 1–2 orders of magnitude more compute (real-benchmark times in App. D.3).

4.2 Real-world experiments

We run our method as the global-positioning stage of GLOMAP (Pan et al., 2024) against its default positioner (GLOMAP-BATA above), first with bundle adjustment and re-triangulation disabled, so the score isolates the positioner, then with the full pipeline, reporting relative-pose AUC at 3/5/10° (per camera pair, the larger of the rotation and translation-direction error). IMC2023 (Google Research, 2023) provides 21 scenes in three families: the small, exhaustively matched *Haiper* and *Heritage* (15–174 images, yet sparse: only 14–24% of Heritage pairs survive geometric verification, ∼ 7 edges per camera) and the large vocabulary-tree *Phototourism* (up to ∼ 900 cameras), all with SIFT features. ETH3D (Schöps et al., 2017) adds 13 DSLR scenes with SuperPoint (DeTone et al., 2018) + LightGlue (Lindenberger et al., 2023) features. EuRoC MAV (Burri et al., 2016) contributes six drone flights in a different regime: a camera moving along a trajectory, matched sequentially, so the view graph is a sparse chain (5.4–11.2 edges per camera) in which many verified pairs carry almost no parallax (6–10% have a baseline under 2 cm, from near-stationary hover).

Table 1: **IMC2023** (Google Research, 2023) (scene count in brackets): mean relative-pose AUC (↑) at $3/5/10^\circ$, replacing GLOMAP's global positioner. *Top:* positioner isolated (bundle adjustment and re-triangulation disabled); *bottom:* full pipeline, three runs per scene. Best results in bold.

	Haiper (3)			Heritage (3)			Phototourism (15)		
Positioner	@3	@5	@10	@3	@5	@10	@3	@5	@10
Positioner only – bundle adjustment and re-triangulation disabled									
Default GLOMAP	24.5	32.6	43.6	16.0	**29.9**	**49.2**	9.5	17.3	30.8
Ours	**24.8**	**33.0**	**43.9**	**16.3**	**29.9**	48.9	**10.6**	**18.8**	**32.6**
Full pipeline – with bundle adjustment and re-triangulation									
Default GLOMAP	55.1	62.8	69.5	72.5	76.6	80.6	71.4	76.9	82.0
Ours	**57.2**	**63.9**	**69.7**	**73.4**	**77.4**	**80.8**	**71.7**	**77.2**	**82.2**

Table 2: **ETH3D** (13 DSLR scenes, SuperPoint+LightGlue features): mean relative-pose AUC (↑) at $3/5/10^\circ$, replacing GLOMAP's global positioner. *Top:* positioner isolated (bundle adjustment and re-triangulation disabled); *bottom:* full pipeline, three runs per scene. Scenes in alphabetical order: (1) Courtyard, (2) Delivery Area, (3) Electro, (4) Facade, (5) Kicker, (6) Meadow, (7) Office, (8) Pipes, (9) Playground, (10) Relief, (11) Relief 2, (12) Terrace, (13) Terrains. Best results in bold.

AUC	Positioner	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	**Mean**
Positioner only – bundle adjustment and re-triangulation disabled															
@3°	Default GLOMAP	17.2	2.2	16.9	49.0	55.2	**0.0**	28.3	**54.2**	15.9	**0.1**	**0.3**	28.7	2.1	20.8
	Ours	**22.9**	**2.3**	**39.0**	**56.3**	**60.8**	**0.0**	**31.5**	**54.2**	**21.1**	**0.1**	**0.3**	**66.3**	**5.7**	**27.7**
@5°	Default GLOMAP	22.8	**4.2**	26.6	56.7	68.0	0.3	38.5	**64.8**	25.8	**0.4**	**0.6**	40.6	4.5	27.2
	Ours	**28.8**	**4.2**	**52.7**	**60.9**	**72.6**	**0.4**	**41.8**	**64.8**	**32.5**	**0.4**	**0.6**	**77.3**	**11.2**	**34.5**
@10°	Default GLOMAP	30.0	**10.0**	41.3	63.6	79.3	**0.9**	53.3	**74.6**	40.5	**1.0**	**1.8**	58.3	10.9	35.8
	Ours	**35.5**	**10.0**	**67.2**	**64.9**	**84.3**	**0.9**	**56.8**	**74.6**	**46.6**	**1.0**	**1.8**	**87.4**	**19.1**	**42.3**
Full pipeline – with bundle adjustment and re-triangulation															
@3°	Default GLOMAP	17.2	30.2	70.7	49.0	76.1	0.0	**50.5**	**72.2**	26.6	0.0	15.8	**84.7**	**88.4**	44.7
	Ours	**38.2**	**30.3**	**70.9**	**62.1**	**80.7**	**35.9**	50.4	**72.2**	**60.8**	**21.5**	**31.5**	84.4	88.3	**55.9**
@5°	Default GLOMAP	22.8	31.9	76.2	56.7	82.3	0.3	**64.0**	**77.1**	39.0	0.0	17.3	**90.8**	**92.9**	50.1
	Ours	**41.2**	**32.0**	**76.4**	**65.0**	**87.6**	**38.7**	63.9	**77.1**	**64.7**	**23.2**	**35.0**	90.6	92.8	**60.6**
@10°	Default GLOMAP	30.0	33.6	81.3	63.6	87.7	1.3	**79.0**	**81.5**	52.4	0.1	18.4	**95.4**	**96.4**	55.4
	Ours	**43.5**	**33.7**	**81.4**	**67.3**	**93.5**	**41.7**	**79.0**	**81.5**	**67.6**	**24.6**	**37.9**	95.3	96.3	**64.9**

Positioner isolated. The top blocks of Tables 1 and 2 show the discrete positioner winning wherever the view graph is well connected. On ETH3D, it wins by a wide margin, $27.7/34.5/42.3$ vs. $20.8/27.2/35.8$: SuperPoint+LightGlue yields dense, low-noise directions on which the global consensus clearly dominates. On Phototourism, it lifts the mean AUC to $10.6/18.8/32.6$ (vs. $9.5/17.3/30.8$), Haiper edges the default ($24.8/33.0/43.9$ vs. $24.5/32.6/43.6$) and Heritage is level ($16.3/29.9/48.9$ vs. $16.0/29.9/49.2$).

With bundle adjustment (BA). We now run the full GLOMAP with BA and re-triangulation (bottom blocks of Tables 1 and 2), replacing only the positioner: the view graph, tracks, re-triangulation and BA are GLOMAP's own, so only the initial positions change. Because GLOMAP's BA is non-deterministic, we run it three times and report the average. The headline is that **our positioner outperforms the default pipeline** on ETH3D (Table 2): $55.9/60.6/64.9$ vs. $44.7/50.1/55.4$, with the largest margins where the default is basin-unstable (*meadow* 38.7 vs. 0.3, *relief* 23.2 vs. 0.0, *playground* 64.7 vs. 39.0, *courtyard* 41.2 vs. 22.8, *kicker* 87.6 vs. 82.3 AUC@5°) and losses only on *office*, *terrace* and *terrains*, none above 0.3. On IMC2023 (Table 1), the margins are smaller but the direction is the same: ours leads all nine columns, $57.2/63.9/69.7$ vs. $55.1/62.8/69.5$ on Haiper, $73.4/77.4/80.8$ vs. $72.5/76.6/80.6$ on Heritage and $71.7/77.2/82.2$ vs. $71.4/76.9/82.0$ on Phototourism, widest at the strict 3° threshold. On these large, well-connected graphs the 3D points carry most of the signal, so the positioner matters less than on ETH3D, where the initialisation decides which basin BA converges to.

Sequential video: EuRoC. Sequential capture changes one thing: the gauge edge. Chosen by two-view support, it lands on consecutive near-stationary frames (support anti-correlates with baseline on every sequence, correlation -0.32 to -0.54), so the scene spans tens of thousands of gauge

Table 3: **EuRoC MAV** (six sequences, stride 10) (Burri et al., 2016): mean relative-pose AUC (↑) at $3/5/10°$, replacing GLOMAP's global positioner. *Top:* positioner isolated (bundle adjustment and re-triangulation disabled); *bottom:* full pipeline, three runs per scene. Best results in bold.

AUC	Positioner	MH_01	MH_02	MH_03	V1_01	V1_02	V1_03	Mean
Positioner only – bundle adjustment and re-triangulation disabled								
@3°	Default GLOMAP	41.5	64.0	41.2	**0.7**	**27.0**	2.9	29.6
	Ours	**43.2**	**68.8**	**42.8**	**0.7**	**27.1**	**3.6**	**31.0**
@5°	Default GLOMAP	63.1	77.8	61.1	6.2	**52.3**	8.0	44.7
	Ours	**64.3**	**80.9**	**63.6**	**6.3**	51.6	**9.3**	**46.0**
@10°	Default GLOMAP	80.7	88.7	78.4	45.8	**75.3**	15.6	64.1
	Ours	**81.5**	**90.4**	**81.2**	**45.9**	74.3	**16.7**	**65.0**
Full pipeline – with bundle adjustment and re-triangulation								
@3°	Default GLOMAP	**48.8**	**75.8**	**44.5**	**0.6**	29.5	**6.2**	34.3
	Ours	**48.9**	**75.8**	**44.6**	**0.6**	**32.2**	**6.2**	**34.7**
@5°	Default GLOMAP	**68.1**	**85.3**	65.9	**6.0**	55.5	**14.3**	49.2
	Ours	**68.1**	**85.4**	**66.1**	**6.0**	**58.0**	**14.3**	**49.6**
@10°	Default GLOMAP	**84.0**	**92.6**	82.5	**46.7**	77.5	22.4	67.6
	Ours	**84.0**	**92.6**	**83.0**	**46.7**	**78.9**	**22.5**	**68.0**

units and the quantisation bound of Prop. 2 is vacuous. Ranking gauge candidates by frame-order span instead restores a well-conditioned frame (median baseline 1.4–7.7 m). Ours wins all three Machine-Hall sequences and the mean, $31.0/46.0/65.0$ vs. $29.6/44.7/64.1$ (Table 3, top block); the two hardest Vicon-room sequences are near-failures for *every* method (0.7 and 2.9 AUC@3° for the default) and carry no signal about positioners. Moreover, our method is deterministic: across three seeds, the default varies by up to 3.6 AUC@3° on one sequence because it starts from random positions, whereas ours is bit-identical on all sequences.

The full pipeline narrows this gap but does not close it. BA absorbs most of the initialisation here: the default's seed sensitivity collapses (its MH_02 spread shrinks below 0.1) and our positioning-stage margin of $+1.4/+1.3/+0.9$ AUC shrinks to $+0.4$ at every threshold, $34.7/49.6/68.0$ vs. $34.3/49.2/67.6$ (Table 3, bottom block). What remains is still a gain: ours matches or beats the default on every sequence at every threshold, tying within 0.1 on MH_01, MH_02, V1_01 and V1_03, and stays ahead on MH_03 (66.1 vs. 65.9) and, most clearly, V1_02 (58.0 vs. 55.5 AUC@5°). Under sequential capture the discrete positioner's advantage thus survives BA, but shrinks far below what it reaches on ETH3D. Sec. 5 returns to why it is hard for translation averaging as such.

5 Conclusions

Global discrete inference could not reach translation averaging, because a camera position offers no bounded label space. We supply one by changing the question: seen from a planted seed, a camera is a direction on S^2, discretised once and reused for every camera and every seed. Each per-seed labeling is a sparse MAP problem that TRW-S optimises globally, robust triangulation of the K rays returns metric positions, and the accompanying analysis (compatibility locus, sparsity, exact depth elimination, maximum-likelihood fusion, quantisation bound) explains why the pieces fit. This buys what no local solver offers: an explicit, joint decision about which edges to trust. On synthetic graphs it yields the lowest error at almost every outlier rate up to a topology-dependent phase transition (80–90% on dense, 70–80% on sparse graphs), halving the error of the state of the art and beating it by 4–6× where contamination is heaviest. Inside GLOMAP it improves or matches the positioning stage on every benchmark, keeps the lead after bundle adjustment (by about ten AUC points on ETH3D), and returns near-identical solutions across runs where the default varies with its random start. Code will be made public.

Limitations: the global solve costs one-to-two orders of magnitude more compute; and sequential forward-dominated motion leaves the along-track baseline weakly determined – a degeneracy of translation averaging itself, afflicting every solver of this problem – yet, used as in GLOMAP, the camera–point constraints supply the missing signal and the full pipeline is unaffected (Table 3).

AI USE STATEMENT

In this work, we used generative AI tools for writing, polishing, and reviewing the text as well as for drafting proofs. Additionally, we used generative AI tools to implement certain parts of the benchmarking scripts as well as for code optimization. We take responsibility for the final content of this work, including text, claims, or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work concerns geometric optimisation for camera-position recovery. All experiments use synthetic data or public benchmarks (ETH3D, Image Matching Challenge 2023, EuRoC MAV) under their published licences; no human subjects, personally identifiable information, or sensitive attributes are involved. Improved structure-from-motion robustness has broadly beneficial uses (mapping, robotics, cultural-heritage digitisation); we are not aware of ethical risks specific to this method beyond those generic to 3D reconstruction.

REPRODUCIBILITY STATEMENT

The method is fully specified in the paper: the model in Secs. 3–3.5, the complete procedure in Algorithm 1, and proofs of all formal claims in Appendix B. Every hyperparameter is listed with its value in Table 4 (Appendix C), and the synthetic data-generation protocol (graph topologies, noise and outlier models, error metric) is given in Sec. 4.1. Real-world experiments use the public GLOMAP pipeline with ETH3D and IMC2023 (SuperPoint+LightGlue and SIFT front-ends) and EuRoC MAV (sequential SIFT matching), with the evaluation protocol described in Sec. 4.2. Source code (C++ core with Python bindings, plus the GLOMAP integration) will be released upon publication.

REFERENCES

Sameer Agarwal, Noah Snavely, Ian Simon, Steven M. Seitz, and Richard Szeliski. Building rome in a day. In *IEEE International Conference on Computer Vision (ICCV)*, pp. 72–79, 2009.

Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, 2001.

Michael Burri, Janosch Nikolic, Pascal Gohl, Thomas Schneider, Joern Rehder, Sammy Omari, Markus W Achtelik, and Roland Siegwart. The EuRoC micro aerial vehicle datasets. *The International Journal of Robotics Research*, 35(10):1157–1163, 2016.

Avishek Chatterjee and Venu Madhav Govindu. Efficient and robust large-scale rotation averaging. In *IEEE International Conference on Computer Vision (ICCV)*, pp. 521–528, 2013.

Avishek Chatterjee and Venu Madhav Govindu. Robust relative rotation averaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4):958–972, 2018.

David Crandall, Andrew Owens, Noah Snavely, and Daniel Huttenlocher. Discrete-continuous optimization for large-scale structure from motion. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3001–3008, 2011.

David J. Crandall, Andrew Owens, Noah Snavely, and Daniel P. Huttenlocher. Sfm with mrfs: Discrete-continuous optimization for large-scale structure from motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(12):2841–2853, 2013.

Zhaopeng Cui and Ping Tan. Global structure-from-motion by similarity averaging. In *IEEE International Conference on Computer Vision (ICCV)*, pp. 864–872, 2015.

Zhaopeng Cui, Nianjuan Jiang, Chengzhou Tang, and Ping Tan. Linear global translation estimation with feature tracks. In *British Machine Vision Conference (BMVC)*, pp. 46.1–46.13, 2015.

Frank Dellaert, David M. Rosen, Jing Wu, Robert Mahony, and Luca Carlone. Shonan rotation averaging: Global optimality by surfing $so(p)^n$ manifolds. In *European Conference on Computer Vision (ECCV)*, volume 12351 of *Lecture Notes in Computer Science*, pp. 292–308. Springer, 2020.

Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Superpoint: Self-supervised interest point detection and description. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 224–236, 2018.

Anders Eriksson, Carl Olsson, Fredrik Kahl, and Tat-Jun Chin. Rotation averaging and strong duality. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 127–135, 2018.

Zhekai Fan, Wanze Li, Jinxin Wang, and Yunpeng Shi. TriP: A triangle puzzle approach to robust translation averaging. arXiv:2605.07143, 2026.

Pedro F. Felzenszwalb and Daniel P. Huttenlocher. Efficient belief propagation for early vision. *International Journal of Computer Vision*, 70(1):41–54, 2006.

Tom Goldstein, Paul Hand, Choongbum Lee, Vladislav Voroninski, and Stefano Soatto. Shapefit and shapekick for robust, scalable structure from motion. In *European Conference on Computer Vision (ECCV)*, pp. 289–304. Springer, 2016.

Google Research. Image matching challenge 2023. Kaggle competition, kaggle.com/competitions/image-matching-challenge-2023, 2023.

Venu Madhav Govindu. Combining two-view constraints for motion estimation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pp. 218–225, 2001.

Venu Madhav Govindu. Lie-algebraic averaging for globally consistent motion estimation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pp. 684–691, 2004.

Richard Hartley, Jochen Trumpf, Yuchao Dai, and Hongdong Li. Rotation averaging. *International Journal of Computer Vision*, 103(3):267–305, 2013.

Peter J. Huber. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1):73–101, 1964.

Nianjuan Jiang, Zhaopeng Cui, and Ping Tan. A global linear method for camera pose registration. In *IEEE International Conference on Computer Vision (ICCV)*, pp. 481–488, 2013.

Vladimir Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(10):1568–1583, 2006.

Shaohan Li, Yunpeng Shi, and Gilad Lerman. Cycle-sync: Robust global camera pose estimation through enhanced cycle-consistent synchronization. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025.

Philipp Lindenberger, Paul-Edouard Sarlin, and Marc Pollefeys. Lightglue: Local feature matching at light speed. In *IEEE International Conference on Computer Vision (ICCV)*, 2023.

Lalit Manam and Venu Madhav Govindu. Correspondence reweighted translation averaging. In *European Conference on Computer Vision (ECCV)*, pp. 56–72. Springer, 2022.

Pierre Moulon, Pascal Monasse, and Renaud Marlet. Global fusion of relative motions for robust, accurate and scalable structure from motion. In *IEEE International Conference on Computer Vision (ICCV)*, pp. 3248–3255, 2013.

Onur Özyeşil and Amit Singer. Robust camera location estimation by convex programming. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2674–2683, 2015.

Onur Özyeşil, Vladislav Voroninski, Ronen Basri, and Amit Singer. A survey of structure from motion. *Acta Numerica*, 26:305–364, 2017.

Linfei Pan, Daniel Barath, Marc Pollefeys, and Johannes L Schönberger. Global structure-from-motion revisited. In *European Conference on Computer Vision (ECCV)*, pp. 58–77, 2024.

Johannes L. Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4104–4113, 2016.

Thomas Schöps, Johannes L. Schönberger, Silvano Galliani, Torsten Sattler, Konrad Schindler, Marc Pollefeys, and Andreas Geiger. A multi-view stereo benchmark with high-resolution images and multi-camera videos. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3260–3269, 2017.

Soumyadip Sengupta, Tal Amir, Meirav Galun, Tom Goldstein, David W. Jacobs, Amit Singer, and Ronen Basri. A new rank constraint on multi-view fundamental matrices, and its application to camera location recovery. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2413–2421, 2017.

Yunpeng Shi and Gilad Lerman. Estimation of camera locations in highly corrupted scenarios: All about that base, no shape trouble. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2868–2876, 2018.

Chitturi Sidhartha and Venu Madhav Govindu. Adaptive annealing for robust averaging. In *European Conference on Computer Vision (ECCV)*, pp. 53–69. Springer, 2024.

Noah Snavely, Steven M. Seitz, and Richard Szeliski. Photo tourism: Exploring photo collections in 3d. *ACM Transactions on Graphics (SIGGRAPH)*, 25(3):835–846, 2006.

Chris Sweeney, Tobias Höllerer, and Matthew Turk. Theia: A fast and scalable structure-from-motion library. In *ACM International Conference on Multimedia (ACM MM)*, pp. 693–696, 2015.

Richard Szeliski, Ramin Zabih, Daniel Scharstein, Olga Veksler, Vladimir Kolmogorov, Aseem Agarwala, Marshall Tappen, and Carsten Rother. A comparative study of energy minimization methods for markov random fields with smoothness-based priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(6):1068–1080, 2008.

Jianyuan Wang, Minghao Chen, Nikita Karaev, Andrea Vedaldi, Christian Rupprecht, and David Novotny. VGGT: Visual geometry grounded transformer. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5294–5306, 2025.

Shuzhe Wang, Vincent Leroy, Yohann Cabon, Boris Chidlovskii, and Jérôme Revaud. DUSt3R: Geometric 3D vision made easy. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20697–20709, 2024.

Kyle Wilson and Noah Snavely. Robust global translations with 1dsfm. In *European Conference on Computer Vision (ECCV)*, pp. 61–75, 2014.

Bingbing Zhuang, Loong-Fah Cheong, and Gim Hee Lee. Baseline desensitizing in translation averaging. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4539–4547, 2018.

Algorithm 1 Translation averaging by spherical direction labeling

Require: view graph $\mathcal{G}$, directions $\{\tilde{\mathbf{t}}_{ij}\}$, confidences $\{w_{ij}\}$; grid $\{\mathbf{v}_a\}_{a=1}^{V}$; K, scales $\mathcal{A}_s$
1: gauge: pick seed edge $(i^\star, j^\star)$, pin $\mathbf{c}_{i^\star}{=}\mathbf{0}$, $\mathbf{c}_{j^\star}{=}\tilde{\mathbf{t}}_{i^\star j^\star}$ (Sec. 3)
2: place seeds $\mathbf{S}_{1..K}$ with cyclic scales; set hard locks (Secs. 3.3, 3.2)
3: build sparse pairwise tensors ψ_{ij} on compatibility tubes (Eq. 9, App. B.3)
4: coarse positions $\hat{\mathbf{c}}_i$ by continuous pre-solve
5: **for** round $= 1, 2$ **do**
6: build unaries $\phi_{k,i}$ from $\hat{\mathbf{c}}_i$ (Eq. 11)
7: **for** $k = 1, \ldots, K$ **in parallel do**
8: discrete decode: TRW-S on (12) $\rightarrow \{\hat{\mathbf{d}}_{k,i}\}$; refinement by IRLS
9: **end for**
10: robust multi-view triangulation $\rightarrow \hat{\mathbf{c}}_i$ (Sec. 3.5)
11: **end for**
12: classify edges by τ_{class}; bridge components; IRLS polish on inliers
13: **return** positions $\{\mathbf{c}_i\}$

A Algorithm

The pseudo-code of the proposed method is shown in Alg. 1.

B Proofs

All results below are elementary geometry, and each proof is written to be read from the picture first. The recurring objects are: a seed $\mathbf{S}_k$ and the candidate direction cells $\mathbf{v}_a, \mathbf{v}_b \in S^2$ in which cameras i and j may appear from it; their unknown depths $r_{k,i}, r_{k,j} > 0$ along those rays and the unknown length $\beta > 0$ of the edge between them; the measured edge direction $\tilde{\mathbf{t}}_{ij}$; the covering radius ρ of the grid and the angular inlier threshold τ; and, for a ray with unit direction $\hat{\mathbf{d}}_k$, the projector $\mathbf{P}_k = \mathbf{I} - \hat{\mathbf{d}}_k \hat{\mathbf{d}}_k^\top$ onto the plane perpendicular to it, so that $\mathbf{P}_k \mathbf{x}$ is the part of $\mathbf{x}$ that is *not* along the ray. Every proof opens with its idea in words and then verifies it in a few named steps.

B.1 Proof of Proposition 1 (compatibility locus)

Proof. Idea. Write both cameras as points on their rays from the seed, $\mathbf{c}_i = \mathbf{S}_k + r_{k,i}\mathbf{v}_a$ and $\mathbf{c}_j = \mathbf{S}_k + r_{k,j}\mathbf{v}_b$, and subtract: the seed cancels and (5) reads $r_{k,j}\,\mathbf{v}_b = r_{k,i}\,\mathbf{v}_a + \beta\,\tilde{\mathbf{t}}_{ij}$. In words: start at the seed, walk $r_{k,i}$ along $\mathbf{v}_a$ to reach camera i, then walk β along the measured edge direction to reach camera j; the direction in which you now see j from the seed is $\mathbf{v}_b$. A sum of two vectors with positive coefficients points into the wedge between them, and that wedge meets the unit sphere in the minor arc joining the two directions. The steps below make this precise; the two extremes of the walk (a very short edge, $\mathbf{v}_b \rightarrow \mathbf{v}_a$; a very long one, $\mathbf{v}_b \rightarrow \tilde{\mathbf{t}}_{ij}$) are the ones described in Sec. 3.2.

Step 1 (every point of the arc admits positive depths). Let $\mathbf{v}_b = (\mathbf{v}_a + s\,\tilde{\mathbf{t}}_{ij})/\|\mathbf{v}_a + s\,\tilde{\mathbf{t}}_{ij}\|$ for some $s > 0$. Reading the depths off this expression, $r_{k,i} = 1$, $\beta = s$ and $r_{k,j} = \|\mathbf{v}_a + s\,\tilde{\mathbf{t}}_{ij}\|$ satisfy (5). All three are positive: the norm could vanish only if $\mathbf{v}_a = -s\,\tilde{\mathbf{t}}_{ij}$, i.e. $s = 1$ and $\mathbf{v}_a = -\tilde{\mathbf{t}}_{ij}$, which the hypothesis excludes.

Step 2 (positive depths force the arc). Conversely, let $r_{k,i}, r_{k,j}, \beta > 0$ satisfy (5). Dividing by $r_{k,i}$,

$$\frac{r_{k,j}}{r_{k,i}}\,\mathbf{v}_b \;=\; \mathbf{v}_a + s\,\tilde{\mathbf{t}}_{ij}, \qquad s = \beta / r_{k,i} > 0,$$

so $\mathbf{v}_b$ is the unit vector along $\mathbf{v}_a + s\,\tilde{\mathbf{t}}_{ij}$ for a positive s, i.e. a point of $\mathcal{A}(\mathbf{v}_a, \tilde{\mathbf{t}}_{ij})$ by definition (6). It remains to see that this set really is the minor arc. The vector $\mathbf{v}_a + s\,\tilde{\mathbf{t}}_{ij}$ never leaves the plane spanned by $\mathbf{v}_a$ and $\tilde{\mathbf{t}}_{ij}$ (a genuine plane, since the two unit vectors are linearly independent by the hypothesis $\mathbf{v}_a \neq \pm\tilde{\mathbf{t}}_{ij}$), and a plane through the origin cuts the sphere in a great circle. Within that plane the vector rotates from $\mathbf{v}_a$ toward $\tilde{\mathbf{t}}_{ij}$ as s grows: its angle from $\mathbf{v}_a$ increases continuously and

strictly, from 0 as $s \to 0$ to $\angle(\mathbf{v}_a, \tilde{\mathbf{t}}_{ij})$ as $s \to \infty$. So as s runs over $(0, \infty)$ the unit vector sweeps the open minor arc from $\mathbf{v}_a$ to $\tilde{\mathbf{t}}_{ij}$, of length $\angle(\mathbf{v}_a, \tilde{\mathbf{t}}_{ij}) < \pi$, exactly once.

Step 3 (why the arc is open, and why only this side of the circle). The two endpoints are not attained: $\mathbf{v}_b = \mathbf{v}_a$ would need $s = 0$, a zero-length edge, and $\mathbf{v}_b = \tilde{\mathbf{t}}_{ij}$ would need $s = \infty$, an infinitely long one; both are excluded by $0 < \beta < \infty$. Negative s sweeps the other half of the same semicircle, from $\mathbf{v}_a$ toward $-\tilde{\mathbf{t}}_{ij}$, and would place camera j *behind* camera i along the measured direction, i.e. satisfy the measurement $-\tilde{\mathbf{t}}_{ij}$ instead; the opposite semicircle, from $\tilde{\mathbf{t}}_{ij}$ through $-\mathbf{v}_a$ to $-\tilde{\mathbf{t}}_{ij}$, is reached for no real s at all. This is the forward branch of Sec. 3.2.

We have shown that positive depths exist exactly for the labels on the open minor arc from $\mathbf{v}_a$ toward $\tilde{\mathbf{t}}_{ij}$. □

B.2 Proof of Corollary 1 (exact depth elimination)

Proof. Idea. With the scale gauged ($r_{k,i} = 1$) the only unknown left in (5) is the depth ratio $\alpha = r_{k,j}$, and the question is whether some $\alpha > 0$ makes the displacement $\alpha\mathbf{v}_b - \mathbf{v}_a$ point along $\tilde{\mathbf{t}}_{ij}$. "Points along" means that the cross product with $\tilde{\mathbf{t}}_{ij}$ vanishes, and that cross product, $\alpha\mathbf{q}_b - \mathbf{q}_a$, is *linear* in α. The question is therefore a one-variable least-squares problem whose residual $E(\alpha) = \|\alpha\mathbf{q}_b - \mathbf{q}_a\|^2$ is a parabola in α, and exact depths exist precisely when that parabola touches zero at a positive α.

Step 1 (the minimiser). Since $\mathbf{q}_b \neq \mathbf{0}$, the leading coefficient $\|\mathbf{q}_b\|^2$ of the parabola is positive, so E is strictly convex and has a unique minimiser, found by setting $E'(\alpha) = 2\,\mathbf{q}_b \cdot (\alpha\mathbf{q}_b - \mathbf{q}_a) = 0$: $\alpha^\star = (\mathbf{q}_b \cdot \mathbf{q}_a)/\|\mathbf{q}_b\|^2$, which is (8).

Step 2 (exact depths give $\alpha^\star > 0$ and $\theta = 0$). Suppose exact positive depths exist. By Prop. 1 there is $s > 0$ with $n\,\mathbf{v}_b = \mathbf{v}_a + s\,\tilde{\mathbf{t}}_{ij}$, where $n = \|\mathbf{v}_a + s\,\tilde{\mathbf{t}}_{ij}\| > 0$. Take the cross product of both sides with $\tilde{\mathbf{t}}_{ij}$: the s term drops out (a vector crossed with itself is zero), leaving $n\,\mathbf{q}_b = \mathbf{q}_a$, i.e. $E(n) = 0$. A parabola that reaches zero attains its minimum there, and the minimiser is unique, so $\alpha^\star = n > 0$ and $E(\alpha^\star) = 0$. Moreover $\alpha^\star\mathbf{v}_b - \mathbf{v}_a = s\,\tilde{\mathbf{t}}_{ij}$ with $s > 0$, so the residual angle $\theta = \angle(\alpha^\star\mathbf{v}_b - \mathbf{v}_a, \tilde{\mathbf{t}}_{ij})$ is 0.

Step 3 ($\alpha^\star > 0$ and $\theta = 0$ give exact depths). Conversely, $\theta = 0$ says that $\mathbf{p} = \alpha^\star\mathbf{v}_b - \mathbf{v}_a$ is a positive multiple of $\tilde{\mathbf{t}}_{ij}$, say $\mathbf{p} = \beta\,\tilde{\mathbf{t}}_{ij}$ with $\beta > 0$. Then $r_{k,i} = 1$, $r_{k,j} = \alpha^\star$ and this β satisfy (5) with all three positive. (The case $\mathbf{p} = \mathbf{0}$, for which θ is undefined, would mean $\alpha^\star\mathbf{v}_b = \mathbf{v}_a$ with both unit vectors, i.e. $\alpha^\star = 1$ and $\mathbf{v}_b = \mathbf{v}_a$: coincident cells, which are excluded as degenerate.)

Step 4 (the excluded case $\mathbf{q}_b = \mathbf{0}$). $\tilde{\mathbf{t}}_{ij} \times \mathbf{v}_b = \mathbf{0}$ means $\mathbf{v}_b = \pm\tilde{\mathbf{t}}_{ij}$. The plus sign is the far endpoint of the arc (6), approached only as the baseline $\beta \to \infty$ and never attained; the minus sign is the backward direction. Neither lies on the open arc, so excluding them loses no compatible pair.

Steps 2 and 3 together show that "exact positive depths exist" and "$\alpha^\star > 0$ with $\theta = 0$" are the same condition. □

B.3 Sparsity of the pairwise potential

Prop. 1 makes the exactly-compatible set a curve, so the gated potential (9) is supported on a thin *tube* around it and only the few cells inside that tube interact. This is a provable property of the model, not merely an implementation trick. The tube is constructed explicitly: the gate of (9) alone would admit far more cells when $\mathbf{v}_a$ approaches $-\tilde{\mathbf{t}}_{ij}$, since then $\mathbf{q}_a \to \mathbf{0}$, $\alpha^\star \to 0^+$ and $\theta \approx 0$ for almost every $\mathbf{v}_b$.

Lemma 1 (Sparsity of the pairwise potential)**.** *Let the grid have V cells of covering radius ρ and near-uniform area (as icosahedral grids do), and construct ψ_{ij} on the geodesic tube of half-width τ around $\mathcal{A}(\mathbf{v}_a, \tilde{\mathbf{t}}_{ij})$. Then the number $\bar{k}$ of cells $\mathbf{v}_b$ interacting with a fixed $\mathbf{v}_a$ satisfies*

$$\bar{k} \;\le\; C\,\frac{V}{4\pi}\Big(2(\tau{+}2\rho)\,L + \pi(\tau{+}2\rho)^2\Big), \qquad L = \angle(\mathbf{v}_a, \tilde{\mathbf{t}}_{ij}) \le \pi, \tag{13}$$

with C a grid-uniformity constant; the active fraction $\bar{k}/V = O(\tau + \rho)$ vanishes as the threshold and grid radius shrink.

Proof. Idea. Count cells by area. Every cell that interacts with $\mathbf{v}_a$ lies inside a thin band around a curve of length L on the sphere; the area of a band is roughly its width times its length, plus two small end caps; and since the cells do not overlap and all have about the same area, the number of cells that fit in the band is at most its area divided by the area of one cell.

Step 1 (which cells are active). By construction $\psi_{ij}(\mathbf{v}_a, \cdot)$ is nonzero only on cells that meet the closed geodesic tube $T_\tau = \{\mathbf{x} \in S^2 : d_g(\mathbf{x}, \mathcal{A}) \le \tau\}$ of half-width τ around the arc $\mathcal{A} = \mathcal{A}(\mathbf{v}_a, \tilde{\mathbf{t}}_{ij})$, whose length is $L = \angle(\mathbf{v}_a, \tilde{\mathbf{t}}_{ij})$; here d_g is geodesic distance.

Step 2 (fatten the band until it contains whole cells). Each cell lies inside the geodesic ball of radius ρ (the covering radius) around its centre. If a cell meets T_τ, some point of it is within τ of the arc, so its centre is within $\tau + \rho$, and then every point of the cell is within $\tau + 2\rho$. Hence all active cells lie entirely inside the wider tube $T_{\tau+2\rho}$, and, being cells of a tiling, they do not overlap.

Step 3 (area of the band). On the unit sphere, the tube of geodesic half-width w around a great-circle arc of length L has area $2L \sin w$ (for the full circle, $L = 2\pi$, this is the familiar area $4\pi \sin w$ of the zone between latitudes $\pm w$), and the two end caps together are at most one geodesic disc of radius w, of area $2\pi(1 - \cos w)$. Using $\sin w \le w$ and $1 - \cos w \le w^2/2$,

$$\operatorname{area}\big(T_w\big) \;\le\; 2L \sin w + 2\pi(1 - \cos w) \;\le\; 2Lw + \pi w^2, \qquad w = \tau + 2\rho.$$

Step 4 (count). Near-uniformity of the icosahedral grid means every cell has area at least $4\pi/(CV)$ for a constant C independent of the subdivision level (empirically $C \lesssim 1.5$). Since the $\bar{k}$ active cells are disjoint and all lie inside $T_{\tau+2\rho}$, their total area is at most that of the band: $\bar{k} \cdot 4\pi/(CV) \le 2L(\tau{+}2\rho) + \pi(\tau{+}2\rho)^2$. Rearranging gives (13). Since $L \le \pi$, the right-hand side is $O\big(V(\tau + \rho)\big)$: the band gets thinner as the threshold and the grid radius shrink, so the active fraction $\bar{k}/V = O(\tau + \rho)$ vanishes.

Numbers. At level 3, $V = 642$ and $\rho \approx 4^\circ \approx 0.07$ rad; with $\tau = 5^\circ \approx 0.087$ rad the band half-width is $w = \tau + 2\rho \approx 0.23$ rad, so $2w \approx 0.46$ and $\pi w^2 \approx 0.16$. Dividing by the mean cell area $4\pi/V$, i.e. taking $C = 1$ (the smallest icosahedral cells are within the factor 1.5 above of the mean), $V/4\pi \approx 51$ and the bound reads $\bar{k} \lesssim 51\,(0.46\,L + 0.16)$: about 10 cells for a short arc (L small) and about 80 for the longest ($L \approx \pi$), bracketing the measured range $17 \le \bar{k} \le 67$. □

B.4 Quantisation error bound

The only systematic error the discretisation introduces is quantisation: each decoded direction is off by at most the covering radius ρ. It propagates to positions in a controlled way.

Proposition 2 (Quantisation error bound)**.** *Let* $\mathbf{c}$ *be the true position,* $r_k = \|\mathbf{c} - \mathbf{S}_k\|$ *its depths, and let the decoded directions satisfy* $\angle(\hat{\mathbf{d}}_k, \mathbf{d}_k) \le \rho$*. If* $\sum_{k\in\mathcal{I}} \hat{\mathbf{P}}_k \succ 0$ *over the inlier rays* $\mathcal{I}$ *of Sec. 3.5, the triangulated position* $\hat{\mathbf{c}}$ *satisfies*

$$\|\hat{\mathbf{c}} - \mathbf{c}\| \;\le\; \frac{\sin\rho \sum_{k\in\mathcal{I}} r_k}{\lambda_{\min}\big(\sum_{k\in\mathcal{I}} \hat{\mathbf{P}}_k\big)}. \tag{14}$$

The bound isolates two levers. The numerator grows with the depths *in gauge units*, so when the camera cluster spans many seed-edge lengths quantisation rather than outliers becomes the binding error source. The denominator is the triangulation conditioning: rays leaving a camera in similar directions drive $\lambda_{\min}$ toward zero, while directionally diverse rays keep it of order $|\mathcal{I}|$ – which is what the seed geometry of Sec. 3.2 is chosen to achieve. The bound thus applies to the rays that survive the consensus, whose diversity sets $\lambda_{\min}$.

Proof. Idea. A ray tilted by an angle of at most ρ misses a point at distance r by at most $r \sin\rho$: lever arm times angle. The triangulated $\hat{\mathbf{c}}$ is the least-squares point at which the decoded rays' miss vectors sum to zero (Step 2 below). At the true $\mathbf{c}$ they do not sum to zero, but each is small, so their sum is small. The matrix $\mathbf{M} = \sum_k \hat{\mathbf{P}}_k$ converts this residual sum into a position offset, and the worst-case conversion rate is $1/\lambda_{\min}(\mathbf{M})$.

Step 1 (how far each decoded ray misses the true point). Write $\hat{\mathbf{P}}_k = \mathbf{I} - \hat{\mathbf{d}}_k \hat{\mathbf{d}}_k^\top$ for the projector of the *decoded* ray. Since $\mathbf{c} - \mathbf{S}_k = r_k \mathbf{d}_k$ with $\mathbf{d}_k$ the true direction, the miss vector of decoded ray k

at the true point is $\hat{\mathbf{P}}_k(\mathbf{c} - \mathbf{S}_k) = r_k \hat{\mathbf{P}}_k \mathbf{d}_k$, and its length satisfies

$$\|\hat{\mathbf{P}}_k \mathbf{d}_k\|^2 = \|\mathbf{d}_k\|^2 - (\hat{\mathbf{d}}_k \cdot \mathbf{d}_k)^2 = 1 - \cos^2\angle(\hat{\mathbf{d}}_k, \mathbf{d}_k) = \sin^2\angle(\hat{\mathbf{d}}_k, \mathbf{d}_k) \;\le\; \sin^2\rho,$$

where the first equality removes the component along $\hat{\mathbf{d}}_k$ and the last step uses $\angle(\hat{\mathbf{d}}_k, \mathbf{d}_k) \le \rho \le \pi/2$. Hence $\|r_k \hat{\mathbf{P}}_k \mathbf{d}_k\| \le r_k \sin\rho$.

Step 2 (compare the estimate with the truth). The estimate solves the normal equations $\mathbf{M}\hat{\mathbf{c}} = \sum_k \hat{\mathbf{P}}_k \mathbf{S}_k$. Subtracting $\mathbf{M}\mathbf{c} = \sum_k \hat{\mathbf{P}}_k \mathbf{c}$ from both sides,

$$\mathbf{M}(\hat{\mathbf{c}} - \mathbf{c}) = \sum_k \hat{\mathbf{P}}_k(\mathbf{S}_k - \mathbf{c}) = -\sum_k r_k\, \hat{\mathbf{P}}_k \mathbf{d}_k,$$

so the position error, seen through $\mathbf{M}$, is minus the sum of the miss vectors of Step 1.

Step 3 (the bound). $\mathbf{M} \succ 0$ is symmetric, so $\|\mathbf{M}^{-1}\| = 1/\lambda_{\min}(\mathbf{M})$. Applying $\mathbf{M}^{-1}$ and then the triangle inequality to the sum,

$$\|\hat{\mathbf{c}} - \mathbf{c}\| \;\le\; \|\mathbf{M}^{-1}\| \Big\| \sum_k r_k \hat{\mathbf{P}}_k \mathbf{d}_k \Big\| \;\le\; \frac{1}{\lambda_{\min}(\mathbf{M})} \sum_k r_k \sin\rho,$$

which is (14).

Step 4 (with IRLS weights). If the triangulation is weighted by $\omega_k \ge 0$, every term above carries a factor ω_k: $\mathbf{M}$ becomes $\sum_k \omega_k \hat{\mathbf{P}}_k$ and the numerator becomes $\sin\rho \sum_k \omega_k r_k$; nothing else changes.

The bound thus separates the two levers discussed above: the depths r_k in gauge units set the size of each miss, and the ray diversity, through $\lambda_{\min}$, sets how strongly the misses are amplified – which is why the seeds of Sec. 3.2 are spread in direction and cycled in distance. □

C Implementation details and hyperparameters

Table 4 lists every hyperparameter with the value used in all experiments (both synthetic and real), unless a section states otherwise. The C++ core implements the sparse compatibility construction (Lemma 1), TRW-S, and the robust triangulation; the K per-seed problems run in parallel threads. TRW-S runs for 50 iterations per decode. The grid-free continuous decode of Sec. 3.4 runs IRLS on the same unary-plus-pairwise objective with directions parameterised continuously on S^2.

Edge confidences. The triplet-closure confidence w_{ij} of (12) scores each edge by the triangles it closes. For every triangle (i, j, k) containing the edge we fit $\tilde{\mathbf{t}}_{ik} \approx a\,\tilde{\mathbf{t}}_{ij} + b\,\tilde{\mathbf{t}}_{jk}$ by least squares; the triangle supports its three edges if $a, b > 0$ (the loop closes with positive baselines) and the residual $r = \|\hat{\mathbf{t}}_{ik} - \tilde{\mathbf{t}}_{ik}\|$ is below 0.2 (about $12°$). An edge's raw score is its number of supporting triangles divided by $1 + 10\,\mathrm{med}(r)$ over them; scores are divided by their median over the supported edges and clipped to $[0.1, 3]$, so an edge closing no triangle keeps a small weight rather than silencing message passing. On mixed camera–point graphs the weights of camera–camera edges are further multiplied by $\gamma = 8$, so that the far more numerous camera–point edges cannot outvote them.

Seed edge and restarts. Gauge candidates are the edges with the highest triplet-closure confidence w_{ij} whose endpoints both have degree at least 3. The top four are each run through the full pipeline, and the run whose triangulated positions leave the most edges within $\tau + \rho$ of their measured direction is kept (ties broken by the median inlier residual). A single corrupted gauge edge therefore spoils one candidate run, not the result.

Soft priors. The coarse positions $\hat{\mathbf{c}}_i$ behind the unaries (11) come from one run of the BATA-style IRLS solver on the raw edges, translated and scaled into the current gauge ($\hat{\mathbf{c}}_{i^\star}{=}\mathbf{0}$, $\|\hat{\mathbf{c}}_{j^\star}\|{=}1$; only these four degrees of freedom are free, so no rotation is needed). Every pre-solved camera enters with the same peak height $w_u = 0.1$. Both choices are deliberate, and we measured them. The *source* is nearly irrelevant: over synthetic graphs at 50–70% outliers, seeding the priors from this pre-solve, from a breadth-first propagation over edges with $w_{uv} \ge 0.3$, or from a solver an order of magnitude more accurate all land within 18% of one another, although the priors themselves differ by more than $30\times$ in accuracy. The *weight* is not: raising it from 0.1 to 0.5 costs a factor of 2.7 at 70% outliers and raising it to 2 a factor of 19, because the prior then overrides the pairwise consensus

that does the outlier rejection; removing it altogether is also worse, since the near-ties of Remark 1 are then broken arbitrarily. A camera that coincides with a gauge seed has no direction from that seed; it and its incident edges are omitted from that seed's MRF, and the other seeds and the hard locks carry their information.

Outer rounds and refinement. The pipeline runs two rounds: after triangulation the unaries are rebuilt from the fused positions and the per-seed problems are solved again. The per-seed refinement of Sec. 3.4 is an IRLS on the same objective with continuous directions. For an edge (i, j) with current directions $\mathbf{d}_i, \mathbf{d}_j$ and the depth ratio $\alpha^\star$ of (8) (clamped to 10^{-3} when non-positive), the direction the edge implies for camera i is the point of the line $\alpha^\star \mathbf{d}_j - \beta\, \tilde{\mathbf{t}}_{ij}$ nearest the current $\mathbf{d}_i$, i.e. $\beta = (\alpha^\star \mathbf{d}_j - \mathbf{d}_i) \cdot \tilde{\mathbf{t}}_{ij}$, normalised (symmetrically $\mathbf{d}_j + \beta'\, \tilde{\mathbf{t}}_{ij}$ with $\beta' = (\alpha^\star \mathbf{d}_i - \mathbf{d}_j) \cdot \tilde{\mathbf{t}}_{ij}$ for camera j). Each camera direction is replaced by the normalised weighted sum of these targets, with weight $\max\{0.02,\ \sigma^4/(\sigma^2 + \theta^2)^2\}\, w_{ij}$ (Geman-McClure on the residual angle θ, $\sigma = \tau + 1.5\rho$), plus the prior direction with weight $w_u = 0.1$; 15 iterations.

Post-processing (Alg. 1, last step). The fused position of each camera is the least-squares point $\arg\min_{\mathbf{c}} \sum_{k \in \mathcal{I}} \|(\mathbf{I} - \hat{\mathbf{d}}_k \hat{\mathbf{d}}_k^\top)(\mathbf{c} - \mathbf{S}_k)\|^2$ over the inlier rays $\mathcal{I}$, a 3×3 solve. With positions in hand, every edge is scored by the angle between $\tilde{\mathbf{t}}_{ij}$ and $\mathbf{c}_j - \mathbf{c}_i$ and marked inlier below $\tau_{\text{class}} = 0.8\,(\tau + 0.5) \approx 27^\circ$ (τ in radians), a deliberately generous gate since the positions are already triangulated. If the inlier subgraph is disconnected, components are bridged by a minimum spanning forest over components: the lowest-residual rejected edge between each pair of components is a candidate, and candidates are re-admitted in increasing residual order until one component remains. The final polish is the BATA solver of Zhuang et al. (2018) on the inlier edges only, with a Huber loss, initialised at the triangulated positions and gauged on the seed edge.

Table 4: Hyperparameters. "Gauge units" are multiples of the seed-edge length (2).

Symbol	Meaning	Value	Where
ℓ	icosahedral subdivision level	3 (V=642, $\rho \approx 4^\circ$)	Sec. 3.1
	at dense N=500	2 (memory bound)	Sec. 4.1
K	number of seeds	8	Sec. 3.2
$\mathcal{A}_s$	cyclic virtual-seed distances	$\{0.5, 1.0, 2.0\}$	Eq. 10
τ	angular inlier threshold	5°	Eq. 9
σ	pairwise Gaussian scale	$\tau/5$	Eq. 9
σ_u	unary prior width	$\tau + 1.5\rho$	Eq. 11
γ	camera-camera boost	8	App. C
	triangulation inlier threshold	$\tau + \rho$ ($\approx 9^\circ$)	Sec. 3.5
	minimal samples / min. inliers	all $\binom{K}{2}$ pairs / 2	Sec. 3.5
	TRW-S iterations	50	Sec. 3.4
τ_{class}	edge-classification threshold	$0.8\,(\tau + 0.5)$, τ in rad	Alg. 1
	seed-edge candidates / selection	4 / most edges within $\tau + \rho$	App. C
	outer rounds	2	Alg. 1
	prior source / weight	BATA pre-solve / 0.1	App. C
	connectivity switch	mean cam-cam degree < 10	Sec. 4.2

D Additional experiments

D.1 Noise sensitivity

Complementing the 50%-outlier noise sweep of Fig. 4, Figure 5 sweeps the per-edge noise $\sigma \in [0, 2]^\circ$ at fixed ρ_{out}=0.3. In the noiseless limit LUD attains *exact recovery* on the dense graph (error 0°, clipped to the display floor in the plot) – a known property of least-unsquared-deviation formulations – but it degrades steeply once noise is present (0.47° already at σ=0.5°, 1.32° at 2°). Among the noise-robust methods, ours is the strongest at every level, degrading smoothly from 0.13° (σ=0) to 0.45° (σ=2°) while GLOMAP-BATA moves from 0.35° to 0.56°; the gap is widest in the low-noise regime, where the 30% outliers – not the noise – are the binding constraint and the discrete decode identifies them more reliably. On sparse graphs the ordering is the same at every noise level: ours degrades from 0.17° (σ=0) to 0.42° (σ=2°) against 0.36° to 0.53° for GLOMAP-BATA, and LUD is again exact at zero noise before rising to 1.00°. The margin narrows from about $2\times$ to about

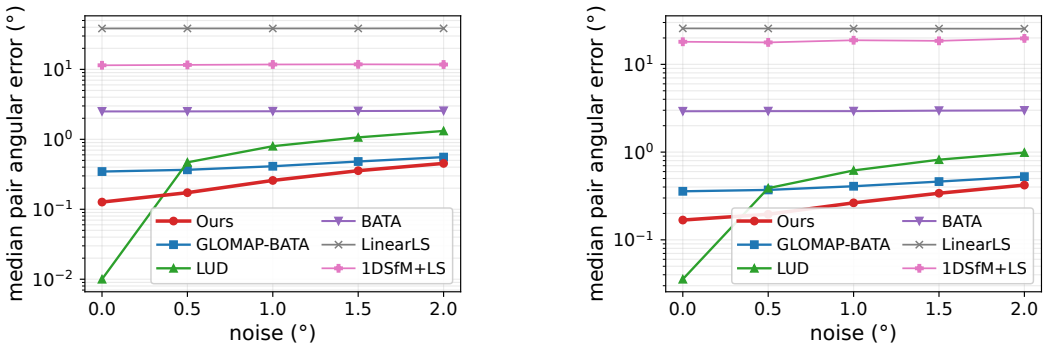


Figure 5: Median pair angular error vs. direction noise (N=100, ρ_{out}=0.3). Left: dense; right: sparse.

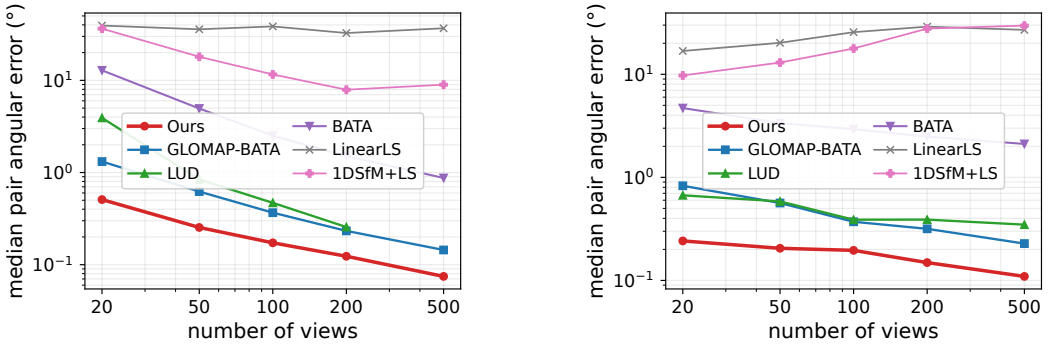


Figure 6: Median pair angular error vs. number of views (ρ_{out}=0.3, noise 0.5°). Left: dense; right: sparse.

1.2× as noise grows: with few edges per view, inlier residuals spread towards the tube half-width τ as σ increases, so the discrete consensus has less to separate and the two methods converge.

D.2 Scaling with problem size

Figure 6 repeats the size sweep of Fig. 4(c) with per-method legends, over $N \in \{20, 50, 100, 200, 500\}$ at fixed ρ_{out}=0.3, noise 0.5°. On dense graphs our advantage is uniform and the error decreases monotonically with size, from 0.51° at N=20 to 0.08° at N=500, a 1.9–2.6× margin over GLOMAP-BATA at every size (LUD is skipped at N=500 for memory). On sparse graphs, where the number of edges grows only linearly with N, the error falls more slowly, from 0.24° to 0.11°, but the margin over GLOMAP-BATA (0.83 → 0.23°) and LUD (0.67 → 0.35°) is preserved at every size, with the largest relative gain on the smallest graphs (N=20: 0.24° vs. 0.83°, 3.4×). At N=20 the k=20 neighbourhood is the whole graph, so the two topologies coincide there and the sparse curve is informative from N=50 upward.

D.3 Runtime

Figure 7 reports wall-clock time for the view-number sweep of Appendix D.2. Every time reported here and in Sec. 4 includes all four gauge-edge restarts and was measured on a 20-thread Intel i9-10900K with 125 GB RAM. On the real benchmarks, GLOMAP's mapper (everything downstream of matching, bundle adjustment included) takes 118 vs. 85 s per ETH3D scene with our positioner vs. the default, 142 vs. 74 s on Heritage, 10.7 vs. 2.7 s on Haiper and 465 vs. 178 s per EuRoC sequence (three runs each; the grid drops to level 2 above 300 cameras). Our method costs 10 s at N=20, 307 s at N=100 and 1103 s at dense N=500 (the level-2 grid keeps the N=500 cost below the N=200 level-3 cost of 1239 s); sparse graphs are 2–4× cheaper at equal N since the compatibility tensor scales linearly in the edge count. The continuous baselines run in seconds – our

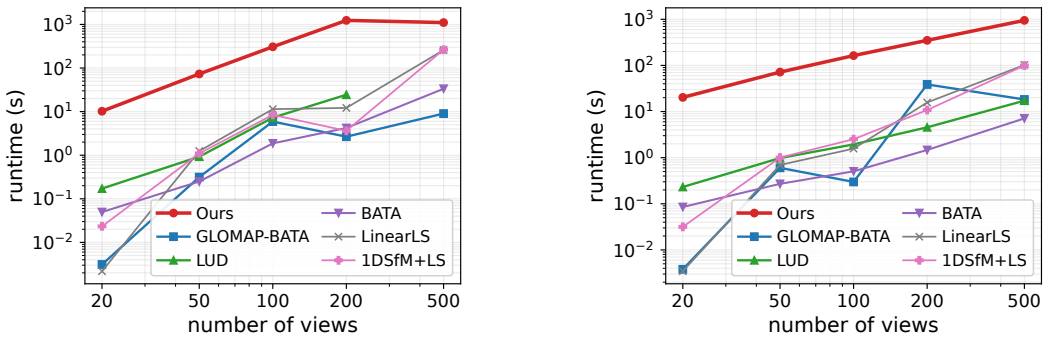


Figure 7: Runtime vs. number of views (ρ_{out}=0.3, noise 0.5°). Left: dense; right: sparse.

accuracy gains are bought with one-to-two orders of magnitude more compute, dominated by the K per-seed compatibility constructions and TRW-S sweeps, both embarrassingly parallel.

D.4 Seed-count and prior-weight ablations

Figure 8 sweeps the two free knobs of the model on one cell (N=50, ER-p=0.5, ρ_{out}=0.5, noise 1°, eight seeds), at the production configuration otherwise.

Seeds (left). Accuracy improves monotonically with K but only slightly, from 0.67° median at K=2 to 0.65° at K=8 (90th percentile 2.00° to 1.84°), while runtime grows only from 57 to 75 s, since the K problems run in parallel. K=2 is not penalised much on this dense cell, where every camera has many edges and the two gauge seeds already triangulate it, but it is fragile by construction: a camera near the seed-edge axis sees both seeds in nearly the same direction, so the two rays are collinear and their least-squares intersection is rank-deficient. With the cyclic per-seed scale assignment of Sec. 3.2, the diagonal seeds $(\mathbf{e}_1 \pm \mathbf{e}_2)/\sqrt{2}$ at K=8 sit at distinct distances from the gauge and contribute genuinely new rays, so we use K=8 throughout: the margin and the insurance are cheap, since the K problems run in parallel.

Prior weight (right). The weight of the soft unary (11) matters far more than where the prior comes from, and the median understates by how much. Both initialisers are optimal at w_u=0.1 (0.646° from the BATA pre-solve, 0.642° from breadth-first propagation) and their curves are indistinguishable at every weight, although the priors themselves differ widely in accuracy. Raising the weight barely moves the median at first (0.70° at w_u=0.5, 0.80° at 1) while the 90th percentile rises from 1.84° to 7.4° and then 20.3°: an over-trusted prior does not so much degrade typical accuracy as produce outright failures on a subset of instances, because it overrides the pairwise consensus that rejects outliers. Removing the prior altogether is also worse (0.77° median, 2.9° p90, dotted line), since the gauge near-ties of Remark 1 are then broken arbitrarily. The prior is therefore load-bearing but must stay weak, which is what w_u=0.1 encodes.

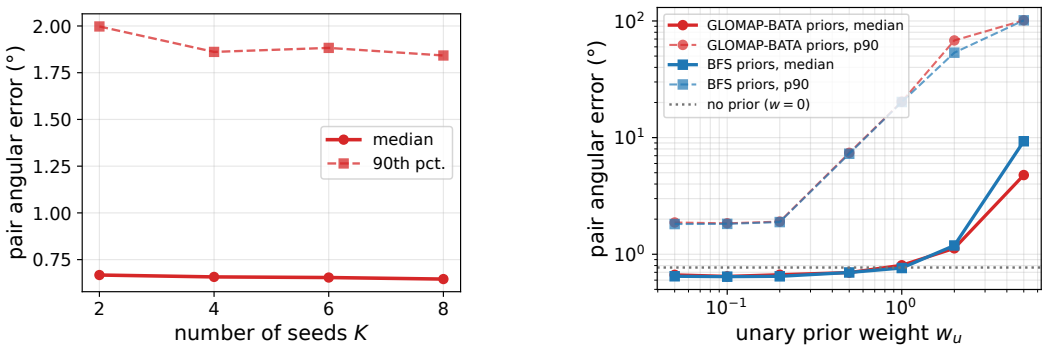


Figure 8: Ablations at N=50, ER-p=0.5, ρ_{out}=0.5, noise 1°. *Left:* number of seeds K. *Right:* unary prior weight, for priors seeded from a GLOMAP-BATA pre-solve and from breadth-first propagation; the dotted line removes the prior, where the two coincide by construction.