RELAXATION-FREE ROBUST VANISHING POINT ESTIMATION IN MANHATTAN WORLD

Anonymous authors
Paper under double-blind review

ABSTRACT

In man-made environments, the parallel edges of a scene form three mutually orthogonal families, known as the Manhattan structure, and their projections meet at three vanishing points that fix the rotation between the camera and the scene. Recovering that rotation from a single calibrated image is difficult, because a line detector returns far more segments than the scene structure actually contains. The two halves of the task are moreover entangled, since a segment can be assigned to a direction only once the directions are known, while estimating them in turn presupposes that assignment. A single criterion, the truncated multi-selection error over the rotation group, settles assignment and estimation together. Our main contribution is MMVP, an iterative algorithm that minimizes this criterion directly, in the original problem dimension and without any relaxation (in the sense of losing tightness). Built on the majorization–minimization principle, the method assembles its surrogate in three stages, freezing the association at the current iterate, majorizing the truncation by an outlier process guided by a homotopy, and dominating the quadratic form over the rotation group by a linear function whose minimizer is available in closed form at a cost linear in the number of lines. We prove that, at every fixed level of the homotopy, the objective values it produces are non-increasing and bounded below, hence convergent, and that the iterates admit limit points, every one of them stationary. A deterministic voting-based initialization completes the pipeline, and no stage of the procedure draws a random sample. Experiments on synthetic scenes and on two real datasets, against geometric and learning-based estimators alike, show that MMVP outperforms them on association F_1 and recall, and that no competing method reaches a higher angular accuracy at any threshold.

1 INTRODUCTION

When a perspective camera images a family of mutually parallel straight lines in a three-dimensional scene, the projections of those lines converge to a single point of the image plane, the *vanishing point* (VP) of that family. It is the image of the point at infinity along the spatial direction the family shares, and it is therefore fixed by that direction and by the camera alone, wherever the individual lines happen to lie [1]. Conversely, a vanishing point detected in a single image measures a three-dimensional scene direction outright, which is what makes it an efficient geometric primitive.

Man-made environments are particularly well suited to this geometric reasoning. Under the *Manhattan world* assumption [2; 3], the dominant straight edges of an indoor or urban scene typically organize into three mutually orthogonal families, the vertical edges of walls and facades together with the two horizontal edge directions of the floor plan. Each family generates a vanishing point, and together they form an orthonormal frame. Recovering this frame is equivalent to determining the rotation between the camera and the scene, a fundamental parameter for visual SLAM [4], camera calibration [5], and scene understanding [6].

A central challenge in this recovery process is the interdependence of its two components. When a line detector is applied to a real image, it produces a large set of segments, only some of which conform to the Manhattan structure. Each segment must be assigned to one of the three families or rejected as an outlier; however, this assignment can only be evaluated once the three directions are known, while the directions themselves can only be estimated after the assignment is determined.

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.

This mutual dependence, more than either component individually, has significantly influenced the development of the literature.

Early approaches address this coupling by alternating between the two components. Latent-assignment schemes [2; 7; 8] iteratively assign lines to directions and re-estimate the directions; while each iteration is computationally light, the assignment step must process every segment. As a result, clutter present in real images is often incorporated into one of the three families instead of being rejected, and soft assignments distribute the influence of outliers across all families instead of suppressing them. Voting schemes bypass the assignment step by accumulating evidence from each line on a discretized Gaussian sphere and identifying dominant peaks [9; 10], though this approach is limited by a fixed resolution. Sampling schemes, such as those in the RANSAC family, hypothesize a Manhattan frame from a minimal subset of lines and retain the hypothesis with the largest consensus [11; 12; 13], with further refinement by multi-model clustering methods like J-Linkage [14] and T-Linkage [15]. These methods tolerate moderate contamination, but a frame constructed from three noisy segments inherits their full noise, and each run may yield a different result. Quasi-global hybrid methods search the rotation space directly for the frame with the largest consensus instead of sampling hypotheses [16; 17; 18]. Their criteria count inliers, so every segment within the threshold contributes equally, regardless of fit quality, and residuals below the threshold do not further refine the estimate. Additionally, the achievable accuracy is constrained by the resolution at which the rotation space is explored, and increasing speed requires coarser resolution.

Recent approaches have addressed the coupling problem using convex optimization. GlobustVP [19] evaluates a frame and an association simultaneously via a *truncated multi-selection error*. The resulting quadratically constrained quadratic program is reformulated by lifting to a higher-dimensional matrix variable, omitting the rank-one constraint, and solving the resulting semidefinite program with a general-purpose conic solver, then striving to recover the original answer via randomization or rounding. This relaxation (in the sense of losing tightness) introduces two structural costs. The first is related to optimality, in that the relaxation may not be tight for all problem dimensions, which can result in suboptimal solutions [20; 21; 22]. The rank-one condition is only recovered by rounding when the relaxed optimum already satisfies it. The second cost is computational, since the lifted variable's size increases with the number of lines, and each interior-point iteration becomes more expensive. To maintain tractability, lines are processed in small random subsets, reintroducing stochasticity that the formulation aimed to eliminate. Furthermore, the runtime depends on the conic solver used, with commercial solvers being fast but open-source alternatives often more than an order of magnitude slower.

Both costs come from relaxing the problem outside the rotation group, which is a property of the solver and not of the objective. We instead minimize the truncated multi-selection error on that group itself, in the dimension in which it is posed, building on the majorization–minimization (MM) principle [23; 24; 25], under which at each iteration a non-convex and non-smooth objective is replaced by a surrogate that dominates it everywhere and touches it at the current iterate, and the minimizer of that surrogate becomes the next iterate. What the principle asks in return is a surrogate at once tight enough for the iterations to make progress and simple enough to be minimized exactly.

Our solver, which we call **MMVP**, constructs its surrogate in three successive stages. First, the inner minimization that selects each line's closest vanishing point is majorized by freezing that selection at the current iterate; since a minimum never exceeds its arguments, the bound is valid everywhere and tight at the iterate, turning the multi-selection error into a sum of quadratic forms, one per direction. Second, the truncation, which confers robustness but flattens the objective far from the solution, is majorized through an outlier process [26], assigning every line a closed-form weight in $[0, 1]$; a homotopy [27; 28] then deforms the truncated loss from a smooth version, in which every line keeps a positive weight, into the exact one, so that a coarse start is tolerated. Third, what remains is minimizing a sum of quadratic forms, one per row of the rotation matrix, over the rotation group; exploiting that every row has unit norm, we dominate each form by a linear function of the rotation, minimized exactly by a singular value decomposition of a small matrix whose size does not depend on the data. Composing the three stages keeps the whole construction within the MM principle, so that the objective is non-increasing at every fixed level of the homotopy, the cost per iteration is linear in the number of lines, and no iterative numerical solver is involved. What separates this from relaxation-based methods is that every iterate is a rotation and every surrogate dominates the exact objective, so the value we report is the objective at a feasible point. The homotopy changes only the loss the surrogate is built for, leaves the feasible set and the variables as they are, and ends at the

exact loss. The problem is non-convex, so the initialization also matters, and we fix it in advance instead of drawing it at random, taking candidate axes from all pairs of lines, completing each to a full frame, and voting on a coarse scan of the one remaining degree of freedom.

We evaluate MMVP against nine competing estimators, geometric and learning-based alike, on synthetic scenes and on two real datasets. It outperforms them in association recovery and in direction estimation together, holds that accuracy as contamination and noise rise, and runs in tens of milliseconds per image. We also measure how near its solutions come to global optimality. In every instance, a posteriori verification shows that the rotation sub-problem at the association we return is solved to global optimality.

Our contributions. The main contributions of this work are summarized as follows.

(i) **A relaxation-free treatment of the robust Manhattan problem.** The truncated multi-selection error is preserved, maintaining the robustness that motivated its adoption, and is minimized directly in the original dimension and on the rotation group, without relaxation (in the sense of losing tightness). No variables are lifted and no constraints are removed, ensuring that the solver retains this property. No relaxed problem yields an optimum lacking the required rank, and no rounding or randomization step is needed.

(ii) **A three-stage majorization whose every iteration is closed form.** All three sources of difficulty, the discrete association, the truncation, and the quadratic dependence on the rotation, admit majorizers that can be minimized exactly, and it is their composition that makes the scheme work. The third majorizer carries the construction, since it is a linear majorizer on the sphere that reduces minimizing a quadratic form over the rotation group to an orthogonal Procrustes problem, and thus to a singular value decomposition of a small fixed-size matrix.

(iii) **Convergence guarantees.** At every fixed level of the homotopy, the sequence of objective values generated by MMVP is non-increasing and bounded below, hence convergent, and the sequence of iterates admits limit points, each of which is a stationary point of the objective being minimized, so the guarantees apply to the sequences themselves.

(iv) **An efficient and solver-free algorithm.** Each iteration requires operations linear in the number of lines, along with a fixed number of small factorizations. The pipeline does not invoke a conic solver, interior-point method, or any other iterative numerical routine. Consequently, the runtime is determined by the method itself, not by third-party software dependencies.

(v) **A deterministic pipeline.** Initialization enumerates line pairs and scans a single angular degree of freedom deterministically, as does the subsequent descent. As a result, the estimate returned for a given image is exactly reproducible, not merely reproducible in distribution, in contrast to sampling-based competitors [12; 13; 14; 15; 29; 30].

(vi) **An a posteriori optimality test that needs no conic solver.** The frame returned by MMVP partitions the lines into three families and a set of outliers. A sufficient condition is derived under which no rotation achieves a lower objective for that partition, thereby settling the geometric component of the problem exactly, in a matter of milliseconds.

(vii) **A broad experimental study.** We compare against state-of-the-art and benchmark methods spanning sampling, multi-model fitting, quasi-global search, convex relaxation, and deep learning, on synthetic scenes and on real datasets, reporting association precision, recall, and F_1, angular accuracy at several thresholds, and runtime, on every dataset.

Notation. Scalars are italic, vectors lowercase bold and matrices uppercase bold. For a matrix $\mathbf{M}$, $\mathbf{M}^\top$ is its transpose, $[\mathbf{M}]_{i,*}$ its i-th row and, when it is symmetric, $\lambda_{\max}(\mathbf{M})$ its largest eigenvalue, with $\mathbf{M} \succeq \mathbf{0}$ marking positive semidefiniteness. Further, $\mathbf{I}_d$ is the $d \times d$ identity matrix, $\langle \mathbf{A}, \mathbf{B} \rangle = \mathrm{tr}(\mathbf{A}^\top \mathbf{B})$ the Frobenius inner product, $\|\cdot\|_2$ the Euclidean norm of a vector, $\otimes$ the Kronecker product, $\mathcal{S}^d$ the set of symmetric $d \times d$ matrices, $\mathbb{1}[\cdot]$ the indicator of an event and $|\mathcal{I}|$ the cardinality of a set $\mathcal{I}$. The unit sphere is $\mathbb{S}^2 = \{\mathbf{d} \in \mathbb{R}^3 \mid \|\mathbf{d}\|_2 = 1\}$, and the orthogonal and rotation groups are $\mathrm{O}(3) = \{\mathbf{R} \in \mathbb{R}^{3\times 3} \mid \mathbf{R}^\top \mathbf{R} = \mathbf{I}_3\}$ and $\mathrm{SO}(3) = \{\mathbf{R} \in \mathrm{O}(3) \mid \det \mathbf{R} = 1\}$.

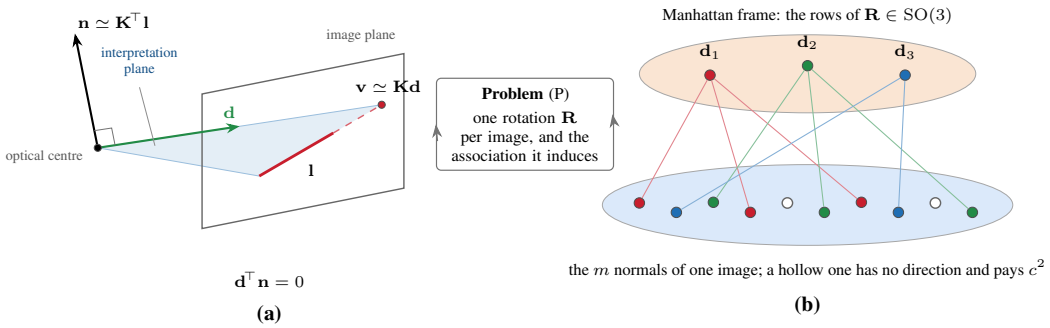

Figure 1: **The geometry of problem** (P)**.** In **(a)** a detected line $\mathbf{l}$ and the optical centre determine an interpretation plane, whose unit normal is $\mathbf{n} \simeq \mathbf{K}^\top \mathbf{l}$ for a calibration matrix $\mathbf{K}$. A scene direction $\mathbf{d}$ can belong to that line only if it lies in the plane, which is what $\mathbf{d}^\top \mathbf{n} = 0$ says, and it is imaged at the vanishing point $\mathbf{v} \simeq \mathbf{K}\mathbf{d}$. In **(b)** the rows $\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3$ of a rotation $\mathbf{R} \in \mathrm{SO}(3)$ are the three directions of a Manhattan frame, and each of the m normals of the image is drawn in the colour of the direction that explains it best, a normal explained by none of them being left hollow and charged c^2, the squared inlier threshold.

2 Problem Formulation

2.1 Vanishing points in a calibrated camera

Let $\mathbf{K} \in \mathbb{R}^{3\times3}$ be the intrinsic calibration matrix of the camera, assumed known, and let us express all spatial quantities in the camera frame. Consider a family of mutually parallel 3D lines with common unit direction $\mathbf{d} \in \mathbb{S}^2$; every line of the family meets the plane at infinity at the same point $(\mathbf{d}^\top, 0)^\top$, which under the perspective projection $\mathbf{x} \simeq \mathbf{K}\,[\,\mathbf{I}_3 \;\; \mathbf{0}\,]\,\mathbf{X}$ (with $\simeq$ denoting equality up to a non-zero scale) is imaged at

$$\mathbf{v} \;\simeq\; \mathbf{K}\,[\,\mathbf{I}_3 \;\; \mathbf{0}\,] \begin{bmatrix} \mathbf{d} \\ 0 \end{bmatrix} \;=\; \mathbf{K}\,\mathbf{d}. \tag{1}$$

The result is precisely the vanishing point $\mathbf{v}$ of the family, the translation component being annihilated by the trailing zero, the algebraic statement that a vanishing point depends on the family's direction and the camera alone, not on the position of its lines.

Consider now an image line $\mathbf{l} \in \mathbb{R}^3$, so an image point $\mathbf{x}$ lies on it iff $\mathbf{x}^\top \mathbf{l} = 0$. The 3D points that project onto $\mathbf{l}$ form a plane through the optical centre, the interpretation plane of the line, with unit normal $\mathbf{n} \in \mathbb{S}^2$ in the camera frame; a short calculation gives $\mathbf{n} \simeq \mathbf{K}^\top \mathbf{l}$, so a calibrated camera converts every detected image line into the unit normal of its interpretation plane.

A line of the family converges to the vanishing point of that family, which in the image plane reads $\mathbf{v}^\top \mathbf{l} = 0$ [1]. Substituting (1) and $\mathbf{n} \simeq \mathbf{K}^\top \mathbf{l}$ into this incidence relation, the calibration matrix cancels,

$$\mathbf{v}^\top \mathbf{l} \;\simeq\; (\mathbf{K}\mathbf{d})^\top \big(\mathbf{K}^{-\top} \mathbf{n}\big) \;=\; \mathbf{d}^\top \mathbf{K}^\top \mathbf{K}^{-\top} \mathbf{n} \;=\; \mathbf{d}^\top \mathbf{n}, \tag{2}$$

and the constraint takes the purely Euclidean form $\mathbf{d}^\top \mathbf{n} = 0$, the single geometric fact on which everything that follows is built, requiring the direction $\mathbf{d}$ to reside within the interpretation plane of the line, orthogonal to its normal, as drawn in Figure 1(a). This is also what makes calibration worth assuming, since it moves the problem from the image plane, where distances are not meaningful, onto the unit sphere, where they are; indeed, $|\mathbf{d}^\top \mathbf{n}|$ represents the sine of the angle between $\mathbf{d}$ and the interpretation plane, and $(\mathbf{d}^\top \mathbf{n})^2$ provides a natural, scale-invariant measure of the incompatibility between a line and a candidate direction.

2.2 The Manhattan frame

Let $\{\mathbf{n}_j\}_{j=1}^m \subset \mathbb{S}^2$ be the normals obtained via $\mathbf{n} \simeq \mathbf{K}^\top \mathbf{l}$ from the m line segments returned by a line detector. Under the Manhattan world assumption, the lines that belong to the scene fall into three families whose directions $\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3 \in \mathbb{S}^2$ are mutually orthogonal, while the rest, produced by clutter, vegetation, reflections or spurious detections, satisfy no such constraint and must be discarded. Collecting the three directions as the rows of a matrix $\mathbf{R} = [\,\mathbf{d}_1 \;\; \mathbf{d}_2 \;\; \mathbf{d}_3\,]^\top \in \mathrm{O}(3)$, their

orthogonality is exactly the statement that $\mathbf{R}$ is orthogonal, and since a vanishing point determines its direction only up to sign, replacing $\mathbf{d}_3$ by $-\mathbf{d}_3$ if necessary restricts the search to $\mathbf{R} \in \mathrm{SO}(3)$ without loss of generality. Estimating the three vanishing points of a Manhattan scene is therefore the same as recovering a single rotation, the one carrying the camera frame onto the scene frame.

2.3 A ROBUST CRITERION FOR JOINT ASSOCIATION AND ESTIMATION

If the association between lines and families were known, estimating $\mathbf{R}$ would become a least-squares problem on $\mathrm{SO}(3)$. However, since the association is unknown, the criterion must jointly evaluate both a frame and an association. Two key observations inform the structure of this criterion. First, it is unnecessary to specify the family membership of each line, because for any candidate frame $\mathbf{R}$, the optimal direction is the one that best explains the line. Thus, line j contributes $\min_{i\in\{1,2,3\}}(\mathbf{d}_i^\top \mathbf{n}_j)^2$, which constitutes a multi-selection error, as the inner minimum implicitly selects the best direction. Second, this error metric must be bounded, because lines that do not belong to any family—common in real images—could otherwise dominate the sum and bias the estimate, a well-known failure mode of least squares when errors are large. To address this, the error is truncated. Fixing an inlier threshold $c > 0$ and capping the contribution of every line at c^2, line j incurs

$$\varepsilon_j(\mathbf{R}) \;=\; \min\Big(\min_{i\in\{1,2,3\}} (\mathbf{d}_i^\top \mathbf{n}_j)^2,\; c^2\Big). \tag{3}$$

A line whose distance to all three directions exceeds c is thus declared an outlier and pays the constant penalty c^2, so that moving the frame slightly no longer changes its contribution. Summing (3) over all lines gives the *truncated multi-selection error*, and the problem we set out to solve is

$$\min_{\mathbf{R}\in\mathrm{SO}(3)} \; F(\mathbf{R}) \;=\; \sum_{j=1}^{m} w_j\, \varepsilon_j(\mathbf{R}), \tag{P}$$

where the $w_j > 0$ are optional per-line confidence weights, taken equal to one unless stated otherwise [19]. A single scalar function of a single variable thus encodes association, estimation and outlier rejection together, which is the situation drawn in Figure 1(b).

Problem (P) remains challenging due to the non-convexity of its feasible set $\mathrm{SO}(3)$, the non-smooth and piecewise nature of its objective as defined by the partitioning of $\mathrm{SO}(3)$ into cells, and the truncation that imparts robustness but also flattens the objective wherever a line is classified as an outlier, resulting in the absence of a gradient far from a solution. This combination of combinatorial selection, non-smooth capping, and a non-convex feasible set has motivated previous work to lift the problem to higher dimensions and relax it in that space [19]. In contrast, the approach presented here addresses each difficulty using a majorizer, thereby demonstrating that lifting is unnecessary.

3 MMVP: A RELAXATION-FREE SOLVER

We construct a majorizer of F at $\mathbf{R}^t$ in three successive stages, building on the majorization–minimization principle [23; 24; 25], since it is precisely this choice that lets each difficulty be dissolved by its own surrogate instead of by a single relaxation of the whole problem, with every surrogate minimized exactly in closed form (for an overview of MM, see Appendix C).

3.1 STAGE ONE: MAJORIZING THE ASSOCIATION

The first difficulty is the inner minimum over the three directions. For each line j we define the selection made at the current iterate,

$$z_j^t \;=\; \operatorname*{arg\,min}_{i\in\{1,2,3\}} \; \big((\mathbf{d}_i^t)^\top \mathbf{n}_j\big)^2, \tag{4}$$

the index of the direction that best explains line j at $\mathbf{R}^t$. Since a minimum never exceeds any of the terms over which it is taken, we obtain, for every $\mathbf{R}$, $\min_{i\in\{1,2,3\}}(\mathbf{d}_i^\top \mathbf{n}_j)^2 \le (\mathbf{d}_{z_j^t}^\top \mathbf{n}_j)^2$, with equality at $\mathbf{R} = \mathbf{R}^t$, by definition (4) of z_j^t. The right-hand side is therefore a valid majorizer of the left-hand side, satisfying both conditions of the MM principle. Its effect is to freeze the association, so that the combinatorial inner minimum disappears, each line is attached to one direction, and the multi-selection error becomes a sum of three ordinary quadratic forms, one per direction.

3.2 Stage two: majorizing the truncation

The second difficulty is the cap. Consider the truncated quadratic $\rho(e) = \min(e, c^2)$ as a function of the squared residual $e \geq 0$. Following the outlier-process construction of [26], we introduce for each line an auxiliary variable $u \in [0, 1]$, interpreted as the degree to which the line is treated as an inlier, and observe that

$$\rho(e) \;=\; \min_{u \in [0,1]} \; \big[\, u\, e \;+\; (1-u)\, c^2 \,\big]. \tag{5}$$

The bracket is affine in u, so its minimum over $[0, 1]$ is attained at an endpoint, e at $u = 1$ when $e < c^2$ and c^2 at $u = 0$ otherwise; the minimizing u is thus the indicator $\mathbb{1}[e \leq c^2]$ of the line being an inlier. This trades the non-smooth cap for a smooth, affine function of e at the price of one extra variable per line, and fixing u at the value that minimizes the bracket for the current residual gives a majorizer of ρ which is tight at that residual.

The substitution is nevertheless only local. Since the truncated loss is flat beyond the threshold, a line that the starting frame places outside it receives $u = 0$ and from then on exerts no influence that could draw the frame back towards it. We therefore minimize, in place of ρ alone, a one-parameter family of losses terminating at ρ, obtained by replacing the penalty $(1 - u)c^2$ with

$$\Phi_\mu(u) \;=\; \frac{\mu\,(1-u)}{\mu + u}\, c^2, \qquad \mu > 0, \tag{6}$$

which defines the losses $\rho_\mu(e) = \min_{u \in [0,1]} [\, u\, e + \Phi_\mu(u) \,]$ of graduated non-convexity [27; 28], indexed by the homotopy parameter μ. As $\mu \to \infty$ we recover $\Phi_\mu(u) \to (1 - u)c^2$, the exact truncated quadratic (5). At the other end, for small μ, the weight (7) below is strictly positive for every line, so no line is switched off and every line keeps pulling the frame towards it. Driving μ from small to large therefore deforms an easy member of the family continuously into the one we actually wish to solve, leaving the feasible set and the variable untouched, and since the exact loss is the last member of the family, the value we finally report is a value of (P) itself.

For a fixed μ, the weight of a line is still available in closed form, so this stage adds a negligible amount of work. Minimizing $u\, e + \Phi_\mu(u)$, which is strictly convex in u, over $[0, 1]$ gives

$$u_j \;=\; \begin{cases} 1, & e_j \leq \frac{\mu}{\mu+1} c^2, \\ \sqrt{\frac{c^2 \mu(\mu+1)}{e_j}} - \mu, & \frac{\mu}{\mu+1} c^2 < e_j < \frac{\mu+1}{\mu} c^2, \\ 0, & e_j \geq \frac{\mu+1}{\mu} c^2, \end{cases} \tag{7}$$

where $e_j = \min_i (\mathbf{d}_i^\top \mathbf{n}_j)^2$ is the residual of line j at the current iterate. The three branches admit a direct reading, in which a line comfortably inside the threshold is a full inlier, one comfortably outside is switched off, and in between the weight decreases smoothly with the residual, the transition band's width governed by μ and shrinking to nothing as the homotopy advances. In practice we set $\mu_0 = c^2/(2e_{\max} - c^2)$, where $e_{\max}$ is the largest residual at initialization, so that no line starts with a vanishing weight, and multiply μ by $\kappa = 2$ every iteration until $\mu \geq 10^6$, beyond which (7) is indistinguishable from the exact truncation.

3.3 Stage three: majorizing the quadratic form over the rotation group

Combining the first two stages, the surrogate to be minimized at iterate t is, up to an additive constant not depending on $\mathbf{R}$, the weighted quadratic

$$G(\mathbf{R} \mid \mathbf{R}^t) \;=\; \sum_{i=1}^{3} \mathbf{d}_i^\top \mathbf{M}_i^t\, \mathbf{d}_i + \text{const}, \qquad \mathbf{M}_i^t \;=\; \sum_{j:\, z_j^t = i} w_j\, u_j^t\, \mathbf{n}_j \mathbf{n}_j^\top \;\succeq\; \mathbf{0}, \tag{8}$$

in which $\mathbf{M}_i^t$ is the weighted scatter matrix of the normals currently attached to direction i. Two of the three difficulties have thus been removed, and the third remains, since minimizing (8) over $\mathrm{SO}(3)$ is still a non-convex problem, and because (8) is quadratic in $\mathbf{R}$, it admits no closed-form minimizer in general.

The way out is to majorize once more, and the geometry of the rotation group is what makes it possible, because every row of a rotation matrix has unit norm, so that on the feasible set a quadratic form may be dominated by a linear one, in exactly the majorize-then-minimize sense recalled in Appendix C.

Lemma 3.1 (Linear majorizer on the sphere). *Let $\mathbf{M} \succeq \mathbf{0}$, let $\lambda = \lambda_{\max}(\mathbf{M})$, and let $\mathbf{d}^t \in \mathbb{S}^2$ be arbitrary. Then, for every $\mathbf{d} \in \mathbb{S}^2$,*

$$\mathbf{d}^\top \mathbf{M} \mathbf{d} \;\le\; 2\,\mathbf{d}^\top (\mathbf{M} - \lambda \mathbf{I}_3)\,\mathbf{d}^t \;+\; \lambda + (\mathbf{d}^t)^\top (\lambda \mathbf{I}_3 - \mathbf{M}) \mathbf{d}^t, \tag{9}$$

with equality at $\mathbf{d} = \mathbf{d}^t$. The right-hand side is thus a valid majorizer of the quadratic form on the unit sphere, and it is affine in $\mathbf{d}$ (for the proof, see Appendix A).

Applying Lemma 3.1 to each of the three rows of $\mathbf{R}$, with $\lambda_i = \lambda_{\max}(\mathbf{M}_i^t)$, and discarding the constants, we obtain the linear surrogate

$$G(\mathbf{R} \mid \mathbf{R}^t) \;\le\; 2\,\langle \mathbf{R},\, \mathbf{G}^t \rangle + \text{const}, \qquad [\mathbf{G}^t]_{i,*} \;=\; \big((\mathbf{M}_i^t - \lambda_i \mathbf{I}_3)\,\mathbf{d}_i^t\big)^\top, \tag{10}$$

affine in $\mathbf{R}$ and tight at $\mathbf{R}^t$. Minimizing an affine function over the rotation group is the orthogonal Procrustes problem, whose global minimizer is known in closed form [31]. Writing the singular value decomposition $-\mathbf{G}^t = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$,

$$\mathbf{R}^{t+1} \;=\; \operatorname*{arg\,min}_{\mathbf{R} \in \mathrm{SO}(3)} \langle \mathbf{R}, \mathbf{G}^t \rangle \;=\; \mathbf{U}\,\mathrm{diag}\big(1,\, 1,\, \det(\mathbf{U}\mathbf{V}^\top)\big)\,\mathbf{V}^\top, \tag{11}$$

the determinant correction ensuring the minimizer is a rotation and not merely an orthogonal matrix. This is the step that makes the whole construction worthwhile, the quadratic subproblem over a non-convex manifold being replaced by an affine one whose exact solution is the singular value decomposition of a matrix whose size does not depend on the data.

Because the linear majorizer of Lemma 3.1 is refreshed at every point, we repeat (10)–(11) a few times at fixed scatter matrices $\mathbf{M}_i^t$, driving the quadratic surrogate (8) further down before the labels and weights are recomputed. We denote by T_{in} the number of such inner iterations, each costing one singular value decomposition of a 3×3 matrix, and set $T_{\text{in}} = 5$ throughout. It is worth mentioning that the three stages are collected in Algorithm 1 (see Appendix D), together with the convergence analysis, the computational complexity, the a posteriori optimality test and the deterministic initialization the descent is run from.

4 Experiments

Setup. All geometric methods are implemented in Python/NumPy and evaluated on the same desktop machine, an Intel Core i7-7700 CPU with 64 GB of memory, to which an NVIDIA RTX 3090 GPU is added for the learning-based baselines alone. We compare MMVP against the four geometric solvers RANSAC [12; 13], J-Linkage [14], T-Linkage [15] and Quasi-VP [17], against the relaxation-based GlobustVP [19], and against the four learning-based methods CONSAC [29], the geometric-prior network [32], PARSAC [30] and the normal-based U-ARE-ME [33]. To keep the comparison fair, every one of them is run in the configuration its own paper describes. What each of these methods does, and the weakness that motivates the present work, are set out in Appendix B. The geometric-prior network [32] and U-ARE-ME read pixels instead of segments, so neither appears in the synthetic comparison. Our own solver uses inlier threshold $c = 0.03$, homotopy factor $\kappa = 2$, $T_{\text{in}} = 5$ inner iterations and $K = 6$ restarts on every dataset, with no per-dataset tuning (further details in Appendix E).

Data. The evaluation utilizes three benchmarks. Synthetic scenes are generated with independently varied outlier ratios and endpoint noise, ensuring exact association by design. HoliCity [34] provides city-scale perspective crops of central London, each annotated with three orthogonal ground-truth vanishing directions. YUD+ [29] offers re-annotated versions of the 102 calibrated urban images from the York Urban dataset, with 31 images annotated with four to eight directions instead of three. Both real datasets are standard benchmarks for this problem, and together they cover scenes in which the Manhattan assumption holds and scenes in which it does not.

Association precision, recall and F_1 score, angular accuracy at 3°, 5° and 10°, and wall-clock runtime are reported across all datasets. Geometric baselines and the proposed solver are executed on CPU, whereas learning-based methods are run on GPU. Consequently, cross-group runtime comparisons provide only an order-of-magnitude estimate. On HoliCity, the runtimes of CONSAC and PARSAC are omitted, as they are evaluated from stored predictions. If a method fails to return a frame, its worst attainable score is included in the averages, ensuring that no method benefits from omitting challenging instances.

Table 1: HoliCity test split. Best per column in **bold**, second best underlined.

Method	Prec. ↑	Rec. ↑	F_1 ↑	AA@3° ↑	AA@5° ↑	AA@10° ↑	time ↓
J-Linkage (2008) [14]	70.0	32.4	39.8	23.4	30.4	38.7	0.003
RANSAC (2012) [12]	80.5	75.5	77.2	55.7	67.1	75.9	0.001
T-Linkage (2014) [15]	70.8	35.5	41.6	24.5	33.0	44.2	0.003
CONSAC (2020) [29]	72.9	81.7	76.5	58.3	65.1	71.1	–
Quasi-VP (2022) [17]	79.8	81.3	79.8	63.0	70.8	78.0	**0.001**
Lin et al. (2022) [32]	**82.4**	74.8	76.8	55.5	61.4	66.5	0.111
PARSAC (2024) [30]	75.9	78.4	76.4	54.6	61.2	67.2	–
U-ARE-ME (2024) [33]	70.3	22.8	31.1	8.8	18.5	43.9	0.033
GlobustVP, MOSEK (2025) [19]	70.3	67.1	68.0	53.1	60.5	66.5	0.571
MMVP (Ours)	82.1	**84.4**	**82.6**	**68.7**	**74.9**	**80.2**	0.036

4.1 Real data: HoliCity

In the HoliCity dataset, the $2{,}490$ test crops from the middle split are used, with lines detected by LSD [35]. As the dataset does not provide per-line labels, each line is labelled by its closest ground-truth direction within the inlier threshold, and every method is scored against these labels. Table 1 presents the complete set of metrics. MMVP achieves the highest performance in recall, F_1, and all three angular thresholds, and is on a par with the best precision in the table.

Two further observations sharpen this ranking. First, the ordering of *association* metrics differs from that of *geometric* metrics. For example, CONSAC ranks second in recall, achieving 81.7 compared to 84.4, but lags by approximately ten points at each angular threshold. This indicates that a method may identify the correct lines overall but may not localize the three directions as precisely. Second, the geometric-prior network [32] demonstrates comparable precision, 82.4 versus 82.1, but ranks fifth at $3°$, so its precision reflects a conservative approach to association, not estimation accuracy. Since both this method and U-ARE-ME process pixels instead of segments, their results are independent of the line detector.

The learning-based methods exhibit a marked divergence in performance in this context. Three of the four methods lose between 9 and 24 points at the $3°$ threshold compared to their results on YUD+ (Appendix H). U-ARE-ME experiences a substantial decrease from 57.3 to 8.8, indicating that its dense normal predictor is the least effective among the four when applied to central-London facades observed at range. In contrast, no competing method reaches a higher angular accuracy than MMVP, which has no learned component, at any threshold on any of the three benchmarks.

GlobustVP fails to return any frame on 415 out of $2{,}490$ crops. These instances are assigned the lowest possible score and kept in the averages (Appendix E).

Figure 2 shows three crops of the test split. The middle crop separates the baselines most sharply, four of the nine recovering no correct association and the best of the rest reaching 75.9, against 88.9 for MMVP at $2.6°$. The bottom crop shows why the two annotations must be read jointly, MMVP scoring 96.1 F_1 with its worst direction $24.0°$ away and every method that recovers the association landing between 23.6 and $25.4°$, which puts the limit in the crop and not in the estimators.

Real data: YUD+. Our second real-data evaluation examines what happens where the Manhattan assumption itself is violated. On YUD+, MMVP outperforms every competing method on precision, recall, F_1 and every angular threshold; the setup, the results, the gallery and the split between Manhattan and non-Manhattan images are in Appendix H.

Further results in the supplementary material. The supplement contains the synthetic study and its generator (Appendix F), the seven metrics at every outlier ratio (Appendix G), the YUD+ study and its Manhattan split (Appendix H), the optimality test (Appendix D.3) and twelve further qualitative panels per dataset (Appendix I).

5 Limitations

Our problem is non-convex, and on problems of this kind a descent method cannot guarantee a global minimizer [36]; what we establish, at every fixed level of the homotopy, is monotone descent of the

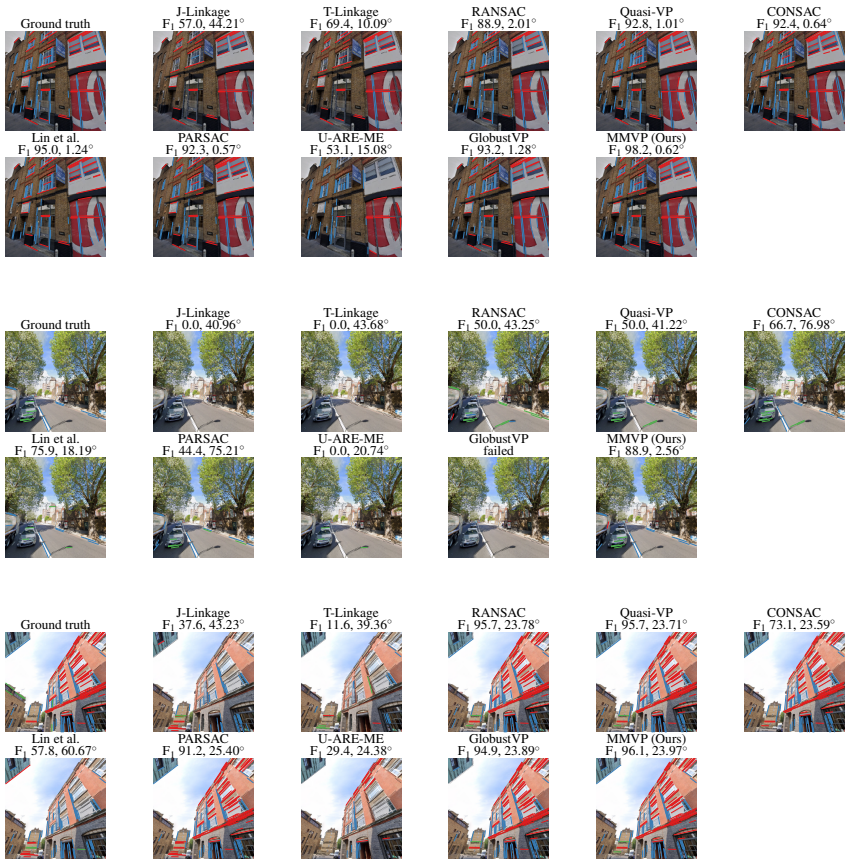

Figure 2: Qualitative results on HoliCity; lines are coloured by association, dotted black for outliers, with F_1 and error above each panel.

objective, convergence of its values, and stationarity of the limit points. Proposition D.2 settles the geometric half of the question, and the combinatorial half remains open.

6 Conclusion

In this study, we addressed the robust estimation of the three orthogonal vanishing points of a Manhattan world from a single calibrated image, where the assignment of lines to directions and the recovery of the directions themselves must be settled together. The truncated multi-selection error evaluates both at once, and our main contribution is a reliable minimization of that error in its original solution space. MMVP reaches this objective in three successive stages, the association, the truncation along a homotopy, and the reduction of the quadratic minimization over $\mathrm{SO}(3)$ to an orthogonal Procrustes problem. Every iteration is therefore closed-form, scales linearly with the number of lines, requires no iterative numerical solver, and leaves every iterate a rotation. We showed that, at every fixed level of the homotopy, the objective values form a non-increasing, bounded and hence convergent sequence, and that the iterates themselves admit stationary limit points. We introduced in addition a deterministic initialization that removes the need for sampling, and derived a sufficient condition—verifiable in milliseconds without a conic solver—under which a rotation is globally optimal for the association MMVP returns. Experiments on synthetic scenes and two real datasets confirm that our approach outperforms existing methods. MMVP leads on association F_1 and recall, and no competing method reaches a higher angular accuracy at any threshold.

Reproducibility Statement

We have made the results of this paper reproducible from the material accompanying the submission. The implementation of MMVP, the synthetic scene generator, the evaluation scripts for all three

benchmarks and the code that produces the tables and figures reported here are provided as a supplementary zip archive submitted together with the paper. The archive carries the complete YUD+ annotations, the HoliCity split we evaluate together with its ground-truth directions, and the stored measurements from which every table regenerates; only the HoliCity photographs are omitted, for their size, and they are distributed by that dataset's own release. The parameter values and the termination criteria of MMVP are stated in Appendix E; the proofs of Lemma 3.1, Theorem D.1 and Proposition D.2 are given in Appendices A, D.1.1 and D.3, respectively.

AI USE STATEMENT

AI-assisted editing was used for language editing and proofreading of the main paper and supplementary materials. The authors take full responsibility for the final content of this paper.

REFERENCES

[1] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2003.

[2] James M Coughlan and Alan L Yuille. Manhattan world: Compass direction from a single image by Bayesian inference. In *ICCV*, 1999.

[3] Julian Straub, Oren Freifeld, Guy Rosman, John J Leonard, and John W Fisher. The Manhattan frame model—Manhattan world inference in the space of surface normals. *IEEE TPAMI*, 40(1):235–249, 2018.

[4] Haoang Li, Yazhou Xing, Ji Zhao, Jean-Charles Bazin, Zhe Liu, and Yun-Hui Liu. Leveraging structural regularity of Atlanta world for monocular SLAM. In *ICRA*, 2019.

[5] Hyunjoon Lee, Eli Shechtman, Jue Wang, and Seungyong Lee. Automatic upright adjustment of photographs with robust camera calibration. *IEEE TPAMI*, 36(5):833–844, 2014.

[6] Alex Flint, David Murray, and Ian Reid. Manhattan scene understanding using monocular, stereo, and 3D features. In *ICCV*, 2011.

[7] Matthew E Antone and Seth Teller. Automatic recovery of relative camera rotations for urban scenes. In *CVPR*, 2000.

[8] Grant Schindler and Frank Dellaert. Atlanta world: An expectation maximization framework for simultaneous low-level edge grouping and camera calibration in complex man-made environments. In *CVPR*, 2004.

[9] Stephen T Barnard. Interpreting perspective images. *Artificial Intelligence*, 21(4):435–462, 1983.

[10] Jean-Philippe Tardif. Non-iterative approach for fast and accurate vanishing point detection. In *ICCV*, 2009.

[11] Martin A Fischler and Robert C Bolles. 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.

[12] Jean-Charles Bazin and Marc Pollefeys. 3-line RANSAC for orthogonal vanishing point detection. In *IROS*, 2012.

[13] Lilian Zhang, Huimin Lu, Xiaoping Hu, and Reinhard Koch. Vanishing point estimation and line classification in a Manhattan world with a unifying camera model. *IJCV*, 117(2):111–130, 2016.

[14] Roberto Toldo and Andrea Fusiello. Robust multiple structures estimation with J-linkage. In *ECCV*, 2008.

[15] Luca Magri and Andrea Fusiello. T-linkage: A continuous relaxation of J-linkage for multi-model fitting. In *CVPR*, 2014.

[16] Haoang Li, Ji Zhao, Jean-Charles Bazin, Wen Chen, Zhe Liu, and Yun-Hui Liu. Quasi-globally optimal and efficient vanishing point estimation in Manhattan world. In *ICCV*, 2019.

[17] Haoang Li, Ji Zhao, Jean-Charles Bazin, and Yun-Hui Liu. Quasi-globally optimal and near/true real-time vanishing point estimation in Manhattan world. *IEEE TPAMI*, 44(3):1503–1518, 2022.

[18] Xiaohu Lu, Jian Yao, Haoang Li, Yahui Liu, and Xiaofeng Zhang. 2-line exhaustive searching for real-time vanishing point estimation in Manhattan world. In *WACV*, 2017.

[19] Bangyan Liao, Zhenjun Zhao, Haoang Li, Yi Zhou, Yingping Zeng, Hao Li, and Peidong Liu. Convex relaxation for robust vanishing point estimation in Manhattan world. In *CVPR*, 2025.

[20] Zhi-Quan Luo, Wing-Kin Ma, Anthony Man-Cho So, Yinyu Ye, and Shuzhong Zhang. Semidefinite relaxation of quadratic optimization problems. *IEEE Signal Processing Magazine*, 27(3):20–34, 2010.

[21] Prabhu Babu, Petre Stoica, and Astha Saini. Fair principal component analysis (PCA): minorization-maximization algorithms for fair PCA, fair robust PCA and fair sparse PCA. *Transactions on Machine Learning Research*, 2025.

[22] Mohammad Mahdi Omati, Prabhu Babu, Petre Stoica, and Arash Amini. A max-min approach to the worst-case class separation problem. *Transactions on Machine Learning Research*, 2025.

[23] Kenneth Lange, David R Hunter, and Ilsoon Yang. Optimization transfer using surrogate objective functions. *J. Computational and Graphical Statistics*, 9(1):1–20, 2000.

[24] David R Hunter and Kenneth Lange. A tutorial on MM algorithms. *The American Statistician*, 58(1): 30–37, 2004.

[25] Ying Sun, Prabhu Babu, and Daniel P Palomar. Majorization-minimization algorithms in signal processing, communications, and machine learning. *IEEE Trans. Signal Process.*, 65(3):794–816, 2017.

[26] Michael J Black and Anand Rangarajan. On the unification of line processes, outlier rejection, and robust statistics with applications in early vision. *IJCV*, 19(1):57–91, 1996.

[27] Andrew Blake and Andrew Zisserman. *Visual Reconstruction*. MIT Press, 1987.

[28] Heng Yang, Pasquale Antonante, Vasileios Tzoumas, and Luca Carlone. Graduated non-convexity for robust spatial perception: From non-minimal solvers to global outlier rejection. *IEEE RA-L*, 5(2):1127–1134, 2020.

[29] Florian Kluger, Eric Brachmann, Hanno Ackermann, Carsten Rother, Michael Ying Yang, and Bodo Rosenhahn. CONSAC: Robust multi-model fitting by conditional sample consensus. In *CVPR*, 2020.

[30] Florian Kluger and Bodo Rosenhahn. PARSAC: Accelerating robust multi-model fitting with parallel sample consensus. In *AAAI*, 2024.

[31] F Landis Markley. Attitude determination using vector observations and the singular value decomposition. *J. Astronautical Sciences*, 36(3):245–258, 1988.

[32] Yancong Lin, Ruben Wiersma, Silvia L Pintea, Klaus Hildebrandt, Elmar Eisemann, and Jan C van Gemert. Deep vanishing point detection: Geometric priors make dataset variations vanish. In *CVPR*, 2022.

[33] Aalok Patwardhan, Callum Rhodes, Gwangbin Bae, and Andrew J Davison. U-ARE-ME: Uncertainty-aware rotation estimation in Manhattan environments, 2024. arXiv:2403.15583.

[34] Yichao Zhou, Jingwei Huang, Xili Dai, Shichen Liu, Linjie Luo, Zhili Chen, and Yi Ma. HoliCity: A city-scale data platform for learning holistic 3D structures, 2020. arXiv:2008.03286.

[35] Rafael Grompone von Gioi, Jérémie Jakubowicz, Jean-Michel Morel, and Gregory Randall. LSD: A fast line segment detector with a false detection control. *IEEE TPAMI*, 32(4):722–732, 2010.

[36] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[37] Jianping Wu, Liang Zhang, Ye Liu, and Ke Chen. Real-time vanishing point detector integrating under-parameterized RANSAC and Hough transform. In *ICCV*, 2021.

[38] Menghua Zhai, Scott Workman, and Nathan Jacobs. Detecting vanishing points using global image context in a non-Manhattan world. In *CVPR*, 2016.

[39] Yichao Zhou, Haozhi Qi, Jingwei Huang, and Yi Ma. NeurVPS: Neural vanishing point scanning via conic convolution. In *NeurIPS*, 2019.

[40] Xin Tong, Xianghua Ying, Yongjie Shi, Ruibin Wang, and Jinfa Yang. Transformer based line segment classifier with image context for real-time vanishing point detection in Manhattan world. In *CVPR*, 2022.

[41] Gwangbin Bae and Andrew J Davison. Rethinking inductive biases for surface normal estimation. In *CVPR*, 2024.

[42] Liangzu Peng, Christian Kümmerle, and René Vidal. On the convergence of IRLS and its variants in outlier-robust estimation. In *CVPR*, 2023.

Supplementary Material

A Proof of Lemma 3.1

Let $\mathbf{M} \succeq \mathbf{0}$, let $\lambda = \lambda_{\max}(\mathbf{M})$, and let $\mathbf{d}, \mathbf{d}^t \in \mathbb{S}^2$. Since λ is the largest eigenvalue of $\mathbf{M}$, we have $\lambda \mathbf{I}_3 - \mathbf{M} \succeq \mathbf{0}$, and therefore the quadratic form of that matrix evaluated at $\mathbf{d} - \mathbf{d}^t$ is non-negative. Expanding it, we obtain

$$0 \;\le\; (\mathbf{d}-\mathbf{d}^t)^\top(\lambda\mathbf{I}_3-\mathbf{M})(\mathbf{d}-\mathbf{d}^t) = \mathbf{d}^\top(\lambda\mathbf{I}_3-\mathbf{M})\mathbf{d} - 2\,\mathbf{d}^\top(\lambda\mathbf{I}_3-\mathbf{M})\mathbf{d}^t + (\mathbf{d}^t)^\top(\lambda\mathbf{I}_3-\mathbf{M})\mathbf{d}^t. \tag{S1}$$

Because $\mathbf{d}$ has unit norm, the first term on the right-hand side equals $\lambda - \mathbf{d}^\top\mathbf{M}\mathbf{d}$. Substituting this identity and rearranging the terms so that the quadratic form in $\mathbf{d}$ stands alone, we arrive at

$$\mathbf{d}^\top\mathbf{M}\mathbf{d} \;\le\; 2\,\mathbf{d}^\top(\mathbf{M}-\lambda\mathbf{I}_3)\,\mathbf{d}^t \;+\; \underbrace{\lambda + (\mathbf{d}^t)^\top(\lambda\mathbf{I}_3-\mathbf{M})\,\mathbf{d}^t}_{\text{const}(\mathbf{d}^t)}, \tag{S2}$$

which is the announced inequality, and whose right-hand side is affine in $\mathbf{d}$. Finally, setting $\mathbf{d} = \mathbf{d}^t$ makes the quadratic form of the first display vanish, so that the inequality holds with equality at that point, and the majorization is tight there. ■

B Related work

Geometric estimators for Manhattan vanishing points. Geometric methods are primarily distinguished by their approach to the line–vanishing-point association. Alternation-based schemes [2; 7; 8] treat this association as a latent variable, updating it alternately with the directions. Although the cost per iteration is negligible, every segment must be assigned to one of the three families. Consequently, image clutter is incorporated into the estimate instead of being excluded, and there is no mechanism to limit the influence of segments inconsistent with all three families. Voting schemes aggregate the contribution of each line on the Gaussian sphere and extract dominant peaks [9; 10]. This approach avoids the need for initialization but restricts attainable accuracy to the resolution of the accumulator. Sampling schemes generate frames from minimal subsets of lines and retain the frame with the largest consensus. These range from standard and orthogonality-aware RANSAC [11; 12; 13; 37] to multi-model clustering methods such as J-Linkage [14; 10] and T-Linkage [15], which cluster lines by their preference sets and do not presuppose the number of models present. These methods tolerate moderate contamination but deteriorate rapidly beyond that point and yield different estimates with each execution. Quasi-global hybrids search the rotation space directly for the frame with the largest consensus, instead of sampling hypotheses [16; 17; 18]. Their criterion counts inliers, so every segment within the threshold is weighted equally, regardless of fit quality. The attainable accuracy is limited by the resolution at which the space is explored, which is often reduced to increase computational speed.

Estimation through convex relaxation. GlobustVP [19] evaluates a frame and an association simultaneously through a truncated multi-selection error. The resulting quadratically constrained quadratic program is lifted to a higher-dimensional matrix variable, the rank-one constraint is omitted, and the semidefinite program that remains is solved with a general-purpose conic solver, one vanishing point at a time and on randomly drawn subsets of lines, after which the method strives to recover the original answer by randomization or rounding. This relaxation (in the sense of losing tightness) may not be tight for every problem instance, and the rank-one condition is recovered only when the relaxed optimum already satisfies it, so the rounded point can be merely feasible and suboptimal [20; 21; 22].

Learning-based estimators. A distinct category of methods trains neural networks to predict vanishing points directly [38; 39; 40]. CONSAC [29] learns a sampling distribution for sequential multi-model fitting, while PARSAC [30] predicts all model instances in parallel. Both approaches function as samplers guided by a learned proposal, so the frame returned by either is determined by a sampling step without an optimality guarantee, and its quality depends on the training data more than on the specific input image. Both methods process line segments and can therefore accept the same input as the proposed solver. The geometric-prior network [32] incorporates a Hough transform that maps pixels to lines and a Gaussian sphere that maps lines to great circles within its architecture. However, its final output is a regression from pixels, influenced by the appearance statistics of its training set and lacking an explicit statement regarding the objective achieved. U-ARE-ME [33]

estimates the Manhattan rotation from dense monocular surface normals [41] using an uncertainty-weighted cost, which ties its accuracy to the underlying normal predictor and limits its performance when those normals are unreliable. Because it reads pixels instead of detected segments, it also provides a reference that does not depend on the line detector. For all four methods, accuracy must be evaluated benchmark by benchmark; none provide guarantees for individual estimates or robustness to changes in dataset, which is specifically assessed in the real-data experiments.

C THE MAJORIZATION–MINIMIZATION FRAMEWORK: AN OVERVIEW

Since our algorithm is built on the majorization–minimization principle, we recall it here in the form in which we use it. Consider the constrained optimization problem

$$\min_{\mathbf{x}\in\mathcal{X}} \; f(\mathbf{x}), \tag{S3}$$

where $\mathbf{x}$ is the decision variable, f is the objective to be minimized and $\mathcal{X}$ denotes the feasible set. At each iteration t, the principle replaces f by a surrogate function $g(\mathbf{x} \mid \mathbf{x}^t)$ that overestimates f everywhere on $\mathcal{X}$ and agrees with it at the current iterate $\mathbf{x}^t$. It places no restriction on f and applies to convex and non-convex, smooth and non-smooth objectives alike. The next iterate is then obtained by minimizing the surrogate instead of the objective,

$$\mathbf{x}^{t+1} \in \operatorname*{arg\,min}_{\mathbf{x}\in\mathcal{X}} \; g(\mathbf{x} \mid \mathbf{x}^t), \tag{S4}$$

and the two operations, construction of the surrogate and its minimization, are repeated until convergence. In the maximization setting the surrogate is required to underestimate the objective instead, and the principle is then referred to as minorization–maximization; the two formulations are equivalent up to a change of sign, and we shall use the acronym MM for both.

For $g(\mathbf{x} \mid \mathbf{x}^t)$ to be a valid majorizer of f at $\mathbf{x}^t$, it must satisfy the two conditions

$$g(\mathbf{x} \mid \mathbf{x}^t) \;\geq\; f(\mathbf{x}) \quad \forall\, \mathbf{x} \in \mathcal{X}, \qquad\qquad g(\mathbf{x}^t \mid \mathbf{x}^t) \;=\; f(\mathbf{x}^t), \tag{S5}$$

usually referred to as the *domination* and the *tangency* conditions, respectively. These two properties alone are enough to guarantee that the objective never increases along the iterations, since

$$f(\mathbf{x}^{t+1}) \;\leq\; g(\mathbf{x}^{t+1} \mid \mathbf{x}^t) \;\leq\; g(\mathbf{x}^t \mid \mathbf{x}^t) \;=\; f(\mathbf{x}^t), \tag{S6}$$

where the first inequality is the domination condition applied at $\mathbf{x}^{t+1}$, the second follows from the fact that $\mathbf{x}^{t+1}$ minimizes the surrogate, and the last equality is the tangency condition. Figure S1 shows the mechanism over three consecutive iterations.

The sandwich (S6) does not require the surrogate to be minimized exactly, since any $\mathbf{x}^{t+1}$ with $g(\mathbf{x}^{t+1} \mid \mathbf{x}^t) \leq g(\mathbf{x}^t \mid \mathbf{x}^t)$ suffices. If, in addition, the objective is bounded below on $\mathcal{X}$, the monotone sequence $\{f(\mathbf{x}^t)\}$ converges, and under mild regularity assumptions the limit points of $\{\mathbf{x}^t\}$ are stationary points of (S3). We refer the reader to [23; 24; 25] for a detailed exposition of the framework and to [42] for its convergence theory.

It is worth mentioning two properties of the framework on which the remainder of the paper rests. First, MM is a design principle and not a particular algorithm, so that what one obtains from it depends entirely on the surrogate one is able to construct, and the whole difficulty of applying it to a new problem lies in finding a majorizer that is tight enough for the iterations to make progress and simple enough to be minimized in closed form. Second, majorizers compose. If $f = f_1 + f_2$ and g_i majorizes f_i at $\mathbf{x}^t$ for $i = 1, 2$, then $g_1 + g_2$ majorizes f at the same point; and if g_1 majorizes f at $\mathbf{x}^t$ and g_2 majorizes g_1 at $\mathbf{x}^t$, then g_2 majorizes f at $\mathbf{x}^t$ as well, since domination and tangency are both preserved by the composition. It is this second property that we exploit in Section 3, where a single majorizer of the truncated multi-selection error is assembled from three successive ones, each of which removes one source of difficulty.

D MMVP: CONVERGENCE, COMPLEXITY, OPTIMALITY AND INITIALIZATION

This appendix gives the complete pseudocode of Algorithm 1, its convergence guarantees, its computational complexity, the a posteriori optimality test applied to its output, and the deterministic initialization it is run from.

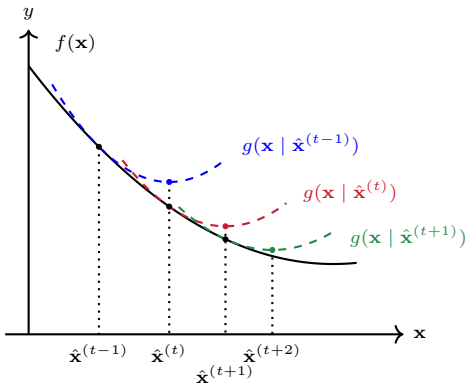

Figure S1: **One iteration of the MM principle.** At the current iterate $\hat{\mathbf{x}}^{(t)}$ a surrogate $g(\cdot \mid \hat{\mathbf{x}}^{(t)})$ is built which lies above the objective f everywhere and touches it at that iterate, the domination and tangency conditions of (S5). The next iterate is the minimizer of the surrogate, not of f, and because the surrogate is tangent from above, the sandwich (S6) makes $f(\hat{\mathbf{x}}^{(t+1)}) \leq f(\hat{\mathbf{x}}^{(t)})$ automatic. The surrogate is then rebuilt at the new iterate, and the objective descends without f itself ever being minimized. The three stages of Section 3 each construct one such surrogate, and their composition is again one.

Algorithm 1 MMVP, the relaxation-free solver built on the MM principle

Require: normals $\{\mathbf{n}_j\}_{j=1}^m$, threshold c, weights $\{w_j\}$, homotopy factor κ=2, inner iterations T_{in}=5, restarts K=6

Generate $\{\mathbf{R}_{(k)}^0\}_{k=1}^K$ by VOTINGINIT (Appendix D.4)
for $k = 1, \ldots, K$ **do**
  Set $\mathbf{R} \leftarrow \mathbf{R}_{(k)}^0$ and initialize μ from the residuals of $\mathbf{R}$
  repeat
    Update the association z_j using (4)
    Update the weights u_j using (7)
    Assemble the scatter matrices $\mathbf{M}_i$ using (8)
    for T_{in} inner iterations **do**
      Update $\mathbf{R}$ using the Procrustes solution (11)
    end for
    if $\mu < 10^6$ **then** $\mu \leftarrow \kappa\mu$
  until $\mu \geq 10^6$ and $F(\mathbf{R})$ is stationary
  $\mathbf{R}_{(k)} \leftarrow \mathbf{R}$
end for
Select $\mathbf{R} \leftarrow \arg\min_k F(\mathbf{R}_{(k)})$ using the exact objective (P)
Recover the labels using (3) and refine $\mathbf{R}$ at fixed labels

Ensure: $\mathbf{R}$ and the line labels

D.1 CONVERGENCE ANALYSIS

We now establish the convergence properties of Algorithm 1. The analysis is carried out at a fixed value of the homotopy parameter, for which we write

$$F_\mu(\mathbf{R}) \;=\; \sum_{j=1}^{m} w_j \, \rho_\mu\big(e_j(\mathbf{R})\big), \qquad e_j(\mathbf{R}) = \min_{i \in \{1,2,3\}} (\mathbf{d}_i^\top \mathbf{n}_j)^2, \tag{S7}$$

the exact objective F of (P) being recovered in the terminal phase of the schedule.

Theorem D.1 (Monotonicity, boundedness and convergence)**.** *Let $\{\mathbf{R}^t\}$ be the sequence of iterates generated by the updates of Algorithm 1 with the homotopy level held fixed at μ. Then the following statements hold.*

(i) Monotonicity. *The objective is non-increasing along the iterations, that is, $F_\mu(\mathbf{R}^{t+1}) \leq F_\mu(\mathbf{R}^t)$ for every t.*

(ii) Boundedness. *The objective is bounded below on the feasible set, since $F_\mu(\mathbf{R}) \geq 0$ for every $\mathbf{R} \in \mathrm{SO}(3)$, and it is bounded above by $c^2 \sum_j w_j$.*

(iii) Convergence of the objective sequence. *Consequently, the sequence $\{F_\mu(\mathbf{R}^t)\}$ converges to a finite limit $F_\mu^\star$.*

(iv) Convergence of the iterates. *The sequence $\{\mathbf{R}^t\}$ admits at least one limit point, every limit point attains the value $F_\mu^\star$, and every limit point is a stationary point of F_μ over $\mathrm{SO}(3)$, or, where F_μ is not differentiable, of a smooth piece of F_μ that is active there.*

In the terminal phase of the schedule, $\mu \geq 10^6$, the same conclusions hold for the exact objective F of (P) *up to a uniform approximation error of at most $10^{-6}c^2$ per line (for the proof, see Appendix D.1.1).*

The proof rests on three facts established below. Each of the three stages produces a majorizer that is tight at the current iterate, the composition of these majorizers is again a majorizer, and (11) minimizes the outermost surrogate globally.

D.1.1 Proof of Theorem D.1

We fix the homotopy level μ throughout and write the objective at that level as

$$F_\mu(\mathbf{R}) = \sum_{j=1}^{m} w_j \, \rho_\mu\big(e_j(\mathbf{R})\big), \qquad e_j(\mathbf{R}) = \min_{i \in \{1,2,3\}} (\mathbf{d}_i^\top \mathbf{n}_j)^2, \tag{S8}$$

where $\rho_\mu(e) = \min_{u \in [0,1]} [\, u\, e + \Phi_\mu(u)\,]$ is the truncated loss at homotopy level μ introduced in Section 3, with the penalty Φ_μ of (6), and whose inner minimizer at a given residual is the closed-form weight (7). At the iterate $\mathbf{R}^t$, with the labels z_j^t of (4) and the weights u_j^t of (7), we define the surrogate

$$G(\mathbf{R} \,|\, \mathbf{R}^t) = \sum_{j=1}^{m} w_j \Big[u_j^t \, \big(\mathbf{d}_{z_j^t}^\top \mathbf{n}_j\big)^2 + \Phi_\mu(u_j^t) \Big]. \tag{S9}$$

The proof of each of the four claims follows.

(i) Monotonicity. We first verify that (S9) is a valid majorizer of F_μ at $\mathbf{R}^t$ in the sense of (S5).

Domination. For any $\mathbf{R} \in \mathrm{SO}(3)$ we may write

$$G(\mathbf{R}|\mathbf{R}^t) \;\geq\; \sum_j w_j \big[u_j^t \, e_j(\mathbf{R}) + \Phi_\mu(u_j^t) \big] \;\geq\; \sum_j w_j \min_{u \in [0,1]} \big[u \, e_j(\mathbf{R}) + \Phi_\mu(u) \big] \;=\; F_\mu(\mathbf{R}), \tag{S10}$$

where the first inequality follows from $(\mathbf{d}_{z_j^t}^\top \mathbf{n}_j)^2 \geq \min_i (\mathbf{d}_i^\top \mathbf{n}_j)^2 = e_j(\mathbf{R})$ together with $u_j^t \geq 0$ and $w_j > 0$, and the second holds because u_j^t is a feasible point of the inner minimization and therefore cannot beat its minimum.

Tangency. At $\mathbf{R} = \mathbf{R}^t$, the index z_j^t attains the inner minimum by its very definition, and u_j^t is the exact minimizer of $u \mapsto u\, e_j(\mathbf{R}^t) + \Phi_\mu(u)$ over $[0,1]$; both inequalities above therefore hold with equality, so that $G(\mathbf{R}^t|\mathbf{R}^t) = F_\mu(\mathbf{R}^t)$.

Descent of the inner iterations. Up to an additive constant, $G(\cdot|\mathbf{R}^t)$ coincides with the quadratic $Q(\mathbf{R}) = \sum_{i=1}^3 \mathbf{d}_i^\top \mathbf{M}_i^t \mathbf{d}_i$ of (8). Applying Lemma 3.1 row by row, we see that for any $\mathbf{R}^{(s)} \in \mathrm{SO}(3)$ the affine function $L_s(\mathbf{R}) = 2\langle \mathbf{R}, \mathbf{G}^{(s)} \rangle + \mathrm{const}$ majorizes Q and is tight at $\mathbf{R}^{(s)}$. Moreover, the Procrustes solution (11) is the *global* minimizer of L_s over $\mathrm{SO}(3)$, since, writing $-\mathbf{G}^{(s)} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$, the maximizer of $\langle \mathbf{R}, -\mathbf{G}^{(s)} \rangle$ over the rotation group is $\mathbf{U}\,\mathrm{diag}(1, 1, \det(\mathbf{U}\mathbf{V}^\top))\,\mathbf{V}^\top$ [31]. Consequently each inner iteration satisfies

$$Q(\mathbf{R}^{(s+1)}) \leq L_s(\mathbf{R}^{(s+1)}) \leq L_s(\mathbf{R}^{(s)}) = Q(\mathbf{R}^{(s)}), \tag{S11}$$

so that the inner loop returns an iterate $\mathbf{R}^{t+1}$ with $G(\mathbf{R}^{t+1}|\mathbf{R}^t) \leq G(\mathbf{R}^t|\mathbf{R}^t)$.

Chaining the three facts, we obtain the MM sandwich

$$F_\mu(\mathbf{R}^{t+1}) \;\leq\; G(\mathbf{R}^{t+1}|\mathbf{R}^t) \;\leq\; G(\mathbf{R}^t|\mathbf{R}^t) \;=\; F_\mu(\mathbf{R}^t), \tag{S12}$$

which is claim (i).

(ii) Boundedness. By construction, $\rho_\mu(e) = \min_{u\in[0,1]}[ue + \Phi_\mu(u)]$ is non-negative, since both $e \geq 0$ and $\Phi_\mu(u) \geq 0$ on $[0,1]$, and it is bounded above by the value $\Phi_\mu(0) = c^2$ of the bracket at $u = 0$. Hence $0 \leq \rho_\mu(e) \leq c^2$ for every residual, and therefore

$$0 \;\leq\; F_\mu(\mathbf{R}) \;\leq\; c^2 \sum_{j=1}^{m} w_j \qquad \text{for every } \mathbf{R} \in \mathrm{SO}(3), \tag{S13}$$

so that the objective is bounded both below and above on the feasible set, uniformly in μ.

(iii) Convergence of the objective sequence. By (i) the real sequence $\{F_\mu(\mathbf{R}^t)\}_{t\geq 0}$ is non-increasing, and by (ii) it is bounded below by zero. The monotone convergence theorem then guarantees that it converges to a finite limit, which we denote by $F_\mu^\star \geq 0$. We stress that this conclusion concerns the sequence of *values* and requires no assumption whatsoever on the iterates themselves.

(iv) Convergence of the iterates. The feasible set $\mathrm{SO}(3)$ is closed and bounded, hence compact, so the sequence $\{\mathbf{R}^t\}$ admits at least one convergent subsequence $\mathbf{R}^{t_k} \to \mathbf{R}^\infty \in \mathrm{SO}(3)$. Since F_μ is continuous on $\mathrm{SO}(3)$, passing to the limit along that subsequence and using (iii) gives

$$F_\mu(\mathbf{R}^\infty) \;=\; \lim_{k\to\infty} F_\mu(\mathbf{R}^{t_k}) \;=\; F_\mu^\star, \tag{S14}$$

so that every limit point attains the same value.

We now turn to stationarity. The labels (4) take finitely many values, so, passing to a further subsequence, we may assume that $z_j^{t_k} = z_j$ for every k and every line j. Each z_j still selects a closest direction at $\mathbf{R}^\infty$, since the inequalities $(\mathbf{d}_{z_j}^\top \mathbf{n}_j)^2 \leq (\mathbf{d}_i^\top \mathbf{n}_j)^2$ that hold along the subsequence persist in the limit. For a frame $\mathbf{R}$, let $\mathbf{G}(\mathbf{R})$ be the matrix of (10) built at $\mathbf{R}$ from these labels and from the weights (7) of the residuals $(\mathbf{d}_{z_j}^\top \mathbf{n}_j)^2$, and let

$$D(\mathbf{R}) \;=\; \big\langle \mathbf{R}, \mathbf{G}(\mathbf{R}) \big\rangle - \min_{\mathbf{R}'\in\mathrm{SO}(3)} \big\langle \mathbf{R}', \mathbf{G}(\mathbf{R}) \big\rangle \;\geq\; 0 \tag{S15}$$

denote the decrease of the linear surrogate achieved by the Procrustes step (11) taken from $\mathbf{R}$. The weights (7) depend continuously on the residual, including at the two ends of the transition band, and the largest eigenvalue depends continuously on the matrix, so D is continuous on $\mathrm{SO}(3)$. The first inner iteration lowers the quadratic surrogate by at least $2D(\mathbf{R}^{t_k})$ and the remaining ones do not raise it, so the sandwich (S12) gives

$$0 \;\leq\; 2D(\mathbf{R}^{t_k}) \;\leq\; F_\mu(\mathbf{R}^{t_k}) - F_\mu(\mathbf{R}^{t_k+1}) \;\xrightarrow[k\to\infty]{}\; 0, \tag{S16}$$

since both values converge to $F_\mu^\star$ by (iii). By continuity, $D(\mathbf{R}^\infty) = \lim_k D(\mathbf{R}^{t_k}) = 0$, that is, $\mathbf{R}^\infty$ minimizes over $\mathrm{SO}(3)$ the linear surrogate built at itself.

This property transfers to F_μ in three steps. Write

$$\mathcal{P}_{\mathbf{R}}(\mathbf{A}) \;=\; \tfrac{1}{2}\big(\mathbf{A}\mathbf{R}^\top - \mathbf{R}\mathbf{A}^\top\big)\mathbf{R} \tag{S17}$$

for the orthogonal projection onto the tangent space $T_{\mathbf{R}}\mathrm{SO}(3) = \{\boldsymbol{\Omega}\mathbf{R} : \boldsymbol{\Omega}^\top = -\boldsymbol{\Omega}\}$, so that a smooth function h has Riemannian gradient $\operatorname{grad} h(\mathbf{R}) = \mathcal{P}_{\mathbf{R}}\big(\nabla h(\mathbf{R})\big)$ on $\mathrm{SO}(3)$. First, the linear surrogate $L(\mathbf{R}) = 2\langle \mathbf{R}, \mathbf{G}(\mathbf{R}^\infty)\rangle + \mathrm{const}$ attains its minimum over $\mathrm{SO}(3)$ at $\mathbf{R}^\infty$, so $\operatorname{grad} L(\mathbf{R}^\infty) = 2\,\mathcal{P}_{\mathbf{R}^\infty}\big(\mathbf{G}(\mathbf{R}^\infty)\big) = \mathbf{0}$. Second, let $Q(\mathbf{R}) = \sum_{i=1}^3 \mathbf{r}_i^\top \mathbf{M}_i \mathbf{r}_i$ be the quadratic surrogate (8) at $\mathbf{R}^\infty$, its scatter matrices $\mathbf{M}_i$ assembled from the labels z_j and the weights u_j of (7) at $\mathbf{R}^\infty$, and let $\lambda_i = \lambda_{\max}(\mathbf{M}_i)$. By Lemma 3.1, applied row by row, L majorizes Q with equality at $\mathbf{R}^\infty$, and the two Euclidean gradients differ by a diagonal multiple of $\mathbf{R}^\infty$, whose projection vanishes,

$$\nabla Q(\mathbf{R}^\infty) - \nabla L(\mathbf{R}^\infty) \;=\; 2\,\mathrm{diag}(\lambda_1,\lambda_2,\lambda_3)\,\mathbf{R}^\infty, \qquad \mathcal{P}_{\mathbf{R}}\big(\mathrm{diag}(\lambda_1,\lambda_2,\lambda_3)\,\mathbf{R}\big) \;=\; \mathbf{0}, \tag{S18}$$

so that $\operatorname{grad} Q(\mathbf{R}^\infty) = \operatorname{grad} L(\mathbf{R}^\infty) = \mathbf{0}$. Third, let $F_\mu^z(\mathbf{R}) = \sum_j w_j\, \rho_\mu\big((\mathbf{d}_{z_j}^\top \mathbf{n}_j)^2\big)$. The loss ρ_μ is continuously differentiable, with derivative equal to the weight (7), because the minimization over u that defines it has a strictly convex objective and hence a unique minimizer. Consequently F_μ^z is continuously differentiable, and since $Q(\mathbf{R}) = \sum_j w_j\, u_j\, (\mathbf{d}_{z_j}^\top \mathbf{n}_j)^2$,

$$\nabla F_\mu^z(\mathbf{R}^\infty) \;=\; \sum_{j=1}^{m} w_j\, u_j\, \nabla\big(\mathbf{d}_{z_j}^\top \mathbf{n}_j\big)^2\Big|_{\mathbf{R}^\infty} \;=\; \nabla Q(\mathbf{R}^\infty), \qquad \operatorname{grad} F_\mu^z(\mathbf{R}^\infty) \;=\; \mathbf{0}. \tag{S19}$$

Moreover, $F_\mu^z \geq F_\mu$ everywhere, because ρ_μ is non-decreasing, and the two coincide at $\mathbf{R}^\infty$. Where F_μ is differentiable at $\mathbf{R}^\infty$, the two functions therefore share their Riemannian gradient there, and $\mathbf{R}^\infty$ satisfies the first-order stationarity condition of F_μ over $\mathrm{SO}(3)$; where a line of positive weight is equidistant from two directions, F_μ has a kink and $\mathbf{R}^\infty$ is a stationary point of the smooth piece F_μ^z, which is active there. This completes the proof of (iv).

The terminal phase. We finally relate the objective at homotopy level μ to the exact one. For every $e \geq 0$ we have $0 \leq \rho(e) - \rho_\mu(e) \leq c^2/(\mu+1)$, since the two losses differ only through the penalty and $\sup_{u\in[0,1]} \left[(1-u)c^2 - \Phi_\mu(u)\right] = c^2/(\mu+1)$. Hence for $\mu \geq 10^6$ the two objectives differ by at most $10^{-6}c^2$ per line, uniformly in $\mathbf{R}$, and the four statements above transfer to the exact objective F of (P) up to that error. ■

Remark on the homotopy. Monotonicity is claimed at a *fixed* μ. Increasing μ between iterations replaces the function being minimized by another member of the family, as happens in every graduated non-convexity scheme [27; 28], so that the guarantee applies to each level of the schedule separately, not across levels. In practice, the exact objective is observed to be non-increasing along essentially the entire trajectory, with occasional small increases immediately after an update of μ in the first few iterations, as illustrated in Appendix D.1.2.

D.1.2 Descent behaviour in practice

Figure S2 presents the exact objective $F(\mathbf{R}^t)$ from (P) across outer iterations for a synthetic instance with $m = 90$ lines and 40% outliers. Each curve corresponds to a different deterministic initialization from Appendix D.4, with five of the $K = 6$ restarts shown. The five curves enter the range between 0.0444 and 0.0520, and all converge to 0.0427. By the tenth iteration, every curve is within 10^{-4} of this value and remains flat thereafter. On this instance, the choice of starting frame changes only the length of the descent path, not the final value. Any increase following an update of μ is negligible at this scale and indistinguishable from the overall descent.

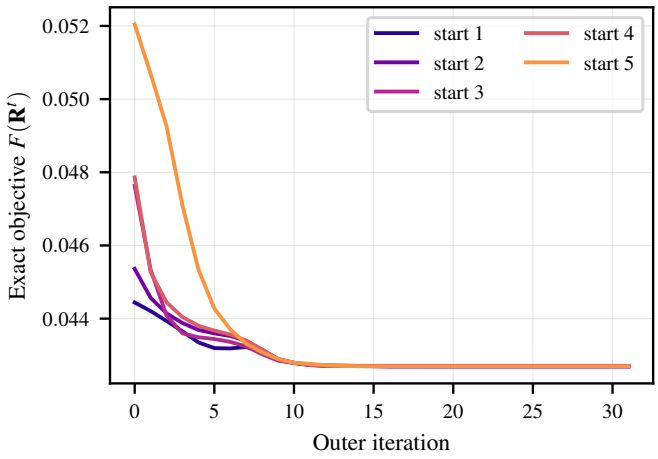

Figure S2: Exact objective $F(\mathbf{R}^t)$ along the outer iterations of MMVP, on one synthetic instance with $m = 90$ lines and 40% outliers, from five of its deterministic starting frames.

D.2 Computational complexity

Each outer iteration evaluates the residuals, labels, and weights of all lines, requiring $O(m)$ operations. The assembly of the three scatter matrices also requires $O(m)$ operations, as each line contributes a single rank-one term. Subsequently, T_{in} inner iterations are performed, each involving the computation of the largest eigenvalue of three 3×3 matrices and a singular value decomposition of a 3×3 matrix. This constitutes a fixed computational cost independent of m. Therefore, the overall cost per iteration is $O(m)$ with minimal constant factors. The number of iterations is determined by the homotopy, not by the problem size. In the experiments, descent typically converges within thirty to forty outer iterations, and the complete solver, including the K restarts and the initialization described in Appendix D.4, completes in a few tens of milliseconds for m up to 135. This contrasts

with a relaxation-based route, in which the variable passed to the conic solver scales with m, and the cost of an interior-point iteration rises even faster, which is why such methods operate on small random subsets of the lines.

D.3 An a posteriori optimality test without a conic solver

Global optimality is a separate question from the descent and stationarity that Theorem D.1 guarantees. A relaxation-based approach does not settle it, since its guarantee depends on a tightness condition that must be verified a posteriori, not assumed [20; 21; 22]. This section shows that a significant part of the question can be answered directly, in a few milliseconds and without a conic solver.

Let $\mathbf{R}^\star \in \mathrm{O}(3)$ be a candidate frame, with rows $\mathbf{r}_i^\star$, and freeze the combinatorial structure it induces, that is, the association $a(\cdot)$ that attaches every inlier line to one of the three directions and the inlier set $\mathcal{I}$ itself. Every outlier then contributes the constant c^2 regardless of the frame, and what remains to be minimized is the smooth problem

$$\min_{\mathbf{R}\mathbf{R}^\top=\mathbf{I}_3} f(\mathbf{R}), \qquad f(\mathbf{R}) = \sum_{i=1}^{3} \mathbf{r}_i^\top \mathbf{M}_i \mathbf{r}_i, \qquad \mathbf{M}_i = \sum_{j\in\mathcal{I},\, a(j)=i} w_j\, \mathbf{n}_j \mathbf{n}_j^\top \;\succeq\; \mathbf{0}. \tag{S20}$$

Stacking the rows into $\mathbf{x} = [\mathbf{r}_1; \mathbf{r}_2; \mathbf{r}_3] \in \mathbb{R}^9$ and writing $\mathbf{Q} = \mathrm{blkdiag}(\mathbf{M}_1, \mathbf{M}_2, \mathbf{M}_3)$, we dualize *both* the row constraints $\mathbf{R}\mathbf{R}^\top = \mathbf{I}_3$, with symmetric multiplier $\mathbf{\Lambda}$, and the algebraically *redundant* column constraints $\mathbf{R}^\top\mathbf{R} = \mathbf{I}_3$, with symmetric multiplier $\mathbf{\Gamma}$, which produces the matrix

$$\mathbf{Z}(\mathbf{\Lambda}, \mathbf{\Gamma}) = \mathbf{Q} - \mathbf{\Lambda} \otimes \mathbf{I}_3 - \mathbf{I}_3 \otimes \mathbf{\Gamma}. \tag{S21}$$

Proposition D.2 (Sufficient condition for global optimality of the frame)**.** *Let $\mathbf{\Gamma} \in \mathcal{S}^3$ and set $\Lambda_{ik} = (\mathbf{r}_k^\star)^\top(\mathbf{M}_i - \mathbf{\Gamma})\mathbf{r}_i^\star$. If (i) the resulting $\mathbf{\Lambda}$ is symmetric and (ii) $\mathbf{Z}(\mathbf{\Lambda}, \mathbf{\Gamma}) \succeq \mathbf{0}$, then $\mathbf{R}^\star$ is a* global *minimizer of* (S20) *over the entire orthogonal group.*

Proof. Let $\mathbf{\Lambda}, \mathbf{\Gamma} \in \mathcal{S}^3$. Writing $\mathbf{x} = [\mathbf{r}_1; \mathbf{r}_2; \mathbf{r}_3] \in \mathbb{R}^9$ for the stacked rows of $\mathbf{R}$ and $\mathbf{c}_j$ for its columns, the Lagrangian of (S20), in which we dualize both the row constraints and the redundant column constraints, is

$$L(\mathbf{x}) = \mathbf{x}^\top \mathbf{Q}\mathbf{x} - \sum_{i,j} \Lambda_{ij}\big(\mathbf{r}_i^\top \mathbf{r}_j - \delta_{ij}\big) - \sum_{j,k} \Gamma_{jk}\big(\mathbf{c}_j^\top \mathbf{c}_k - \delta_{jk}\big) = \mathbf{x}^\top \mathbf{Z}\mathbf{x} + \mathrm{tr}\,\mathbf{\Lambda} + \mathrm{tr}\,\mathbf{\Gamma}, \tag{S22}$$

where we have used the two identities $\sum_{i,j} \Lambda_{ij} \mathbf{r}_i^\top \mathbf{r}_j = \mathbf{x}^\top(\mathbf{\Lambda} \otimes \mathbf{I}_3)\mathbf{x}$ and $\sum_{j,k} \Gamma_{jk} \mathbf{c}_j^\top \mathbf{c}_k = \sum_i \mathbf{r}_i^\top \mathbf{\Gamma} \mathbf{r}_i = \mathbf{x}^\top(\mathbf{I}_3 \otimes \mathbf{\Gamma})\mathbf{x}$, together with the definition (S21) of $\mathbf{Z}$. At any feasible $\mathbf{R}$ both constraint terms vanish, so that $f(\mathbf{R}) = L(\mathbf{x})$; combined with hypothesis (ii), this gives the lower bound

$$f(\mathbf{R}) \;=\; \mathbf{x}^\top \mathbf{Z}\mathbf{x} + \mathrm{tr}\,\mathbf{\Lambda} + \mathrm{tr}\,\mathbf{\Gamma} \;\geq\; \mathrm{tr}\,\mathbf{\Lambda} + \mathrm{tr}\,\mathbf{\Gamma} \tag{S23}$$

for *every* feasible $\mathbf{R}$, that is, over the whole orthogonal group.

The candidate $\mathbf{R}^\star$ attains this bound. Indeed, the i-th block of $\mathbf{Z}\mathbf{x}^\star$ is $(\mathbf{M}_i - \mathbf{\Gamma})\mathbf{r}_i^\star - \sum_j \Lambda_{ij}\mathbf{r}_j^\star$. Since $\mathbf{R}^\star$ is orthogonal, its rows form an orthonormal basis of $\mathbb{R}^3$, so it suffices to check that every component of that block along $\mathbf{r}_k^\star$ vanishes; and indeed this component equals $(\mathbf{r}_k^\star)^\top(\mathbf{M}_i - \mathbf{\Gamma})\mathbf{r}_i^\star - \Lambda_{ik}$, which is zero by the definition of $\mathbf{\Lambda}$ in the statement. Hence $\mathbf{Z}\mathbf{x}^\star = \mathbf{0}$ and

$$f(\mathbf{R}^\star) = (\mathbf{x}^\star)^\top \mathbf{Z}\mathbf{x}^\star + \mathrm{tr}\,\mathbf{\Lambda} + \mathrm{tr}\,\mathbf{\Gamma} = \mathrm{tr}\,\mathbf{\Lambda} + \mathrm{tr}\,\mathbf{\Gamma}, \tag{S24}$$

which meets the lower bound (S23). The duality gap is therefore zero and $\mathbf{R}^\star$ is a global minimizer of (S20) over $\mathrm{O}(3)$. Finally, hypothesis (i) is what makes the construction legitimate, since the Lagrangian above requires $\mathbf{\Lambda}$ to be symmetric; written out, it reads $(\mathbf{r}_k^\star)^\top \mathbf{M}_i \mathbf{r}_i^\star = (\mathbf{r}_i^\star)^\top \mathbf{M}_k \mathbf{r}_k^\star$, the contributions of $\mathbf{\Gamma}$ cancelling by its own symmetry, so that (i) is a property of the candidate alone and coincides with the first-order optimality condition of (S20). ■

Condition (i) pertains solely to the candidate and aligns with the first-order optimality condition of (S20). The redundant column constraints are essential for ensuring the tightness of the test. Since $\mathbf{\Lambda}$ has been eliminated, $\mathbf{Z}$ is affine in the six free parameters of $\mathbf{\Gamma}$. Consequently, verifying the condition reduces to a six-dimensional concave maximization that can be solved without a conic solver, with a median computational cost of approximately five milliseconds.

Finally, we clarify the scope of the test. It addresses the geometric component of the problem, in that it demonstrates that no rotation improves upon $\mathbf{R}^\star$ for the association and inlier set induced by $\mathbf{R}^\star$. The test succeeds on every instance we examined (Appendix F), so for the association MMVP returns, the rotation sub-problem is solved to global optimality, not left at a poorly converged point or a suboptimal local minimum.

D.4 Deterministic initialization

The homotopy widens the region from which the descent reaches the lowest plateau, and because the problem is non-convex, the starting frame also matters. MMVP therefore starts from deterministic frames and retains several candidates, which is affordable because each descent takes only milliseconds.

Initialization is performed in two steps. First, every pair of lines (a, b) hypothesizes a direction $\mathbf{v} = \mathbf{n}_a \times \mathbf{n}_b / \|\mathbf{n}_a \times \mathbf{n}_b\|_2$, as a direction compatible with two lines must lie in both interpretation planes. All $\binom{m}{2}$ pairs are enumerated, or a fixed subset is used when m is large. Each hypothesis is evaluated by its inlier support $|\{j : |\mathbf{v}^\top \mathbf{n}_j| < c\}|$, and the most angularly distinct hypotheses are retained as candidate dominant axes. In the second step, each candidate axis leaves one degree of freedom undetermined, specifically a rotation by an angle α about the axis. The parameter α is scanned over a coarse grid of $[0, \pi/2)$, the three angles with the largest support are retained, and additional frames are formed from pairs of voted directions that are nearly orthogonal. The resulting set of approximately twelve candidate frames is ranked using the exact objective (P), and the descent of Algorithm 1 is initiated from the best $K = 6$ candidates. The precise parameters of this procedure are provided in Appendix E.

Voting over line pairs is inherently a coarse method, with its accuracy limited by the resolution of the grid, which also constrains searches such as Quasi-VP [17]. In this context, voting is used solely to provide an initial frame. Subsequent iterations jointly re-estimate all three directions using soft weights and all lines simultaneously, achieving a level of accuracy unattainable by voting alone.

Because (P) is non-convex, the initialization also matters, as for any method in this class, and two evaluations quantify its role. As we saw in Appendix D.1.2, the exact objective tracked from five deterministic frames produced by the described procedure, on an instance with 40% contamination, reaches the same lowest plateau from all five. Appendix F adds a complementary bound, in that an exhaustive search over the rotation space finds no lower value than MMVP on 93% of the instances, and the test of Proposition D.2 confirms, for every instance examined, that the rotation sub-problem at the returned association is solved to global optimality. Alternative initializations therefore offer little additional benefit, and the K candidates are retained because each descent is cheap, not because any particular candidate requires fine-tuning.

E Implementation details

MMVP. We use $T_{\text{in}} = 5$ inner Procrustes iterations per outer iteration, with an early exit when the decrease of the surrogate falls below 10^{-12}, a homotopy factor $\kappa = 2$, and we terminate each descent once $\mu \geq 10^6$ and the exact objective changes by less than 10^{-10}. The deterministic initialization enumerates all $\binom{m}{2}$ line pairs when $\binom{m}{2} \leq 4000$, and a fixed subset of 4000 pairs otherwise; the best angularly distinct axis hypotheses, taken at least 10° apart, are completed to full frames by scanning 180 angles on $[0, \pi/2)$ and keeping the three non-adjacent peaks of largest support per axis, to which we add the frames built from pairs of voted directions that are close to orthogonal. The resulting set is ranked by the exact objective and the descent is run from the best $K = 6$ candidates. After the candidate of lowest objective has been selected, the labels are extracted through (3) and a final Manhattan refinement is applied, in which the labels are held fixed and a weighted least-squares problem over $\mathrm{SO}(3)$ is solved by the same Procrustes iterations.

Baselines. Every competing method is run in the configuration described in its own paper.

Failure scoring. If a method raises an error, or exhausts its budget without producing a frame, the trial is scored $F_1 = 0$ and 90° of angular error per direction, and the elapsed time is recorded as it stands. No trial is removed from the averages, since dropping the failures would favour precisely the methods that give up most often.

One row of Table 1 shows what that choice is worth. GlobustVP returns no frame at all on 415 of the 2,490 HoliCity crops, and each of those is scored at 90° per direction with $P = R = \mathrm{F}_1 = 0$.

F SYNTHETIC DATA EXPERIMENTS

We generate Manhattan scenes whose outlier ratio and measurement noise are controlled independently. A gravity-aligned frame with uniform yaw and realistic camera tilt (pitch $\sim \mathcal{N}(0, 10°)$, roll $\sim \mathcal{N}(0, 5°)$) is viewed by a 640×480 pinhole camera with a 60° field of view; inlier segments are axis-aligned 3D segments projected into the image with Gaussian endpoint noise, and outliers are uniformly random image segments. The default setting has $m = 90$ lines and an endpoint noise of $\sigma = 1$ px, with 50 Monte-Carlo trials per setting, reduced to 25 for the secondary studies. Since the generator knows which line belongs to which family, the ground-truth association is exact here, and precision, recall and F_1 are computed against the true labels.

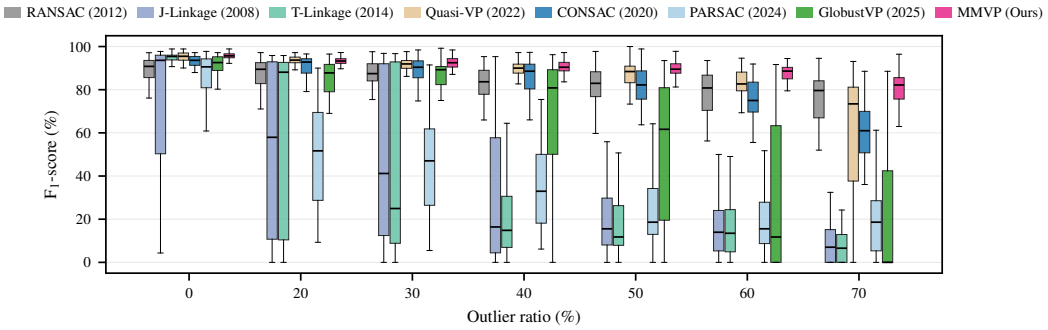

Figure S3: Robustness to outliers on synthetic scenes, as grouped boxplots of the F_1-score of the eight methods evaluable on line input, each labelled with its publication year.

Robustness to outliers. Figure S3 varies the outlier ratio from 0% to 70%. MMVP attains the highest mean F_1 at 0% contamination and at every ratio from 30% upwards, with 92.5 at 30%, 90.6 at 40%, 89.6 at 50%, 87.6 at 60% and 77.6 at 70%, against 91.7, 89.4, 86.6 and 78.2 for Quasi-VP, the nearest competitor up to 60%, and 74.3 for RANSAC, which takes that place at 70%. At 20% contamination MMVP and Quasi-VP are on a par, within 0.2 of each other, and from 30% onwards the gap widens steadily in our favour, which is precisely the regime the truncated objective is designed for. The two linkage methods stay below 24 from 50% on. Our median worst-direction error remains below 0.9° over the whole range, from 0.32° at 0% to 0.82° at 70%. The boxes also show that MMVP is the most stable method across trials, since its box is the shortest at every ratio other than 30%, where it is on a par with that of Quasi-VP.

GlobustVP degrades quickly as contamination grows, its F_1 falling to 67.1 at 40% outliers and to 20.6 at 70%, because the randomized conic solves it runs on small subsets of lines become unreliable once those subsets are contaminated.

Noise and scalability. Raising the endpoint noise to $\sigma = 5$ px at 30% outliers isolates what a non-minimal estimator buys. MMVP is the most stable method in the comparison and still reaches a mean F_1 of 50.0 at $\sigma = 5$, ahead of Quasi-VP at 47.8 and RANSAC at 47.4, because a minimal hypothesis inherits the full endpoint noise of the few segments that generate it, whereas the weighted rotation step of Section 3 averages that noise down over every line the association admits. GlobustVP falls to 15.0 and the linkage methods below 12 over the same range. The scalability study of Figure S4 makes the efficiency gap concrete, in that our runtime holds between 0.030 and 0.038 s as m grows from 60 to 135, with F_1 flat at 93 and leading the comparison across the range, Quasi-VP drawing level at a single size, whereas the relaxation-based pipeline needs between 1.0 and 1.4 s, since, as anticipated in Appendix D.2, our iteration count is set by the homotopy, not by m, and each iteration is linear in m with very small constants.

Distance to the global optimum. The problem is non-convex, so how often MMVP reaches the global minimum of (P) is established empirically. As a reference, we use an exhaustive search over the rotation space, 2×10^5 samples polished by minimization of the exact loss, on 100 held-out

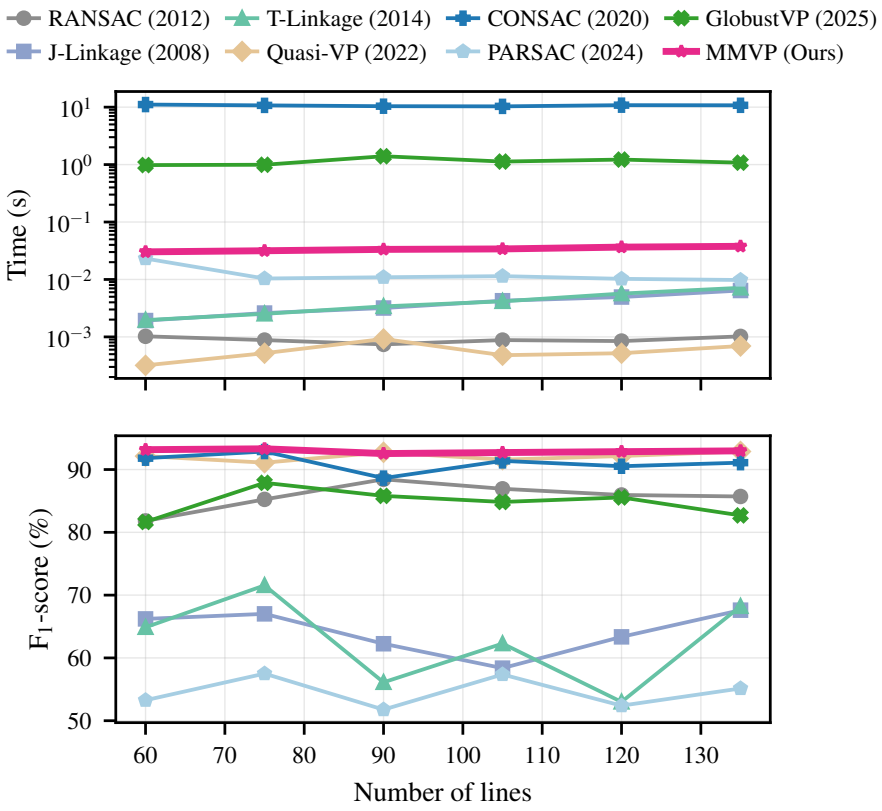

Figure S4: Runtime and accuracy as the number of lines grows from 60 to 135 at 20% outliers.

instances split evenly over four outlier ratios. Table S1 collects the outcome. Against this reference, MMVP attains the global optimum on 93% of the instances, on all of them at 0% and 20% outliers, on 96% at 40% and on 76% at 60%; the reference costs about 0.5 s per instance, more than fifteen times the runtime of MMVP.

Table S1: The optimality study on 100 synthetic instances with $m = 90$ lines and $\sigma = 1$ px endpoint noise, 25 trials at each outlier ratio. The first row compares MMVP with the exhaustive reference; the second reports how often the a posteriori test of Proposition D.2 succeeds.

Outlier ratio	0%	20%	40%	60%	all
Match with the exhaustive reference (%)	100.0	100.0	96.0	76.0	93.0
Optimality test succeeds (%)	100.0	100.0	100.0	100.0	100.0
Median cost of the test (ms)	4.0	5.0	5.0	5.0	5.0
Median MMVP runtime (ms)	30.0	31.5	34.5	32.5	32.0
Median runtime of the reference (s)	0.53	0.50	0.53	0.51	0.52

Verifying the geometry, and localizing the residual gap. The optimality test described in Proposition D.2 renders the 93% figure considerably more informative. When applied to the same set of 100 instances, as reported in the second row of Table S1, the test succeeds in all cases, at a median cost of 5.0 ms and without a conic solver. The rotation sub-problem is therefore solved exactly on every instance, and wherever the exhaustive reference reaches a lower value, the difference lies in the association or the inlier set, not in a poorly converged rotation or a suboptimal local minimum.

The synthetic result under the real-data metric suite. The preceding studies report F_1 and angular error, which clarify performance trends. Table S2 presents the same seven metrics used for both real datasets, evaluated at a representative operating point of $m = 90$ lines, $\sigma = 1$ px endpoint

Table S2: Synthetic Manhattan scenes at 30% outliers. Best per column in **bold**, second best underlined.

Method	Prec. ↑	Rec. ↑	F_1 ↑	AA@3° ↑	AA@5° ↑	AA@10° ↑	time ↓
J-Linkage (2008) [14]	77.1	46.2	50.7	46.0	49.3	54.7	0.003
RANSAC (2012) [12]	92.5	82.8	87.1	89.3	94.0	97.3	0.001
T-Linkage (2014) [15]	76.5	43.0	46.2	42.0	46.7	50.7	0.003
CONSAC (2020) [29]	89.4	86.7	87.9	90.0	90.7	92.0	10.814
Quasi-VP (2022) [17]	92.1	91.4	91.7	98.7	**100.0**	**100.0**	**0.001**
PARSAC (2024) [30]	64.0	36.4	44.9	24.0	30.0	38.0	0.010
GlobustVP, MOSEK (2025) [19]	89.5	77.8	81.8	81.3	85.3	90.0	1.590
MMVP (Ours)	**93.3**	**91.7**	**92.5**	**100.0**	**100.0**	**100.0**	0.034

noise and 30% outliers, over 50 seeded trials. This approach enables direct comparison between synthetic and real evidence. MMVP ranks first in precision, recall, F_1, and across all three angular thresholds. At 5° and 10°, Quasi-VP also achieves saturation, resulting in a shared first place. The faster baselines each reach a lower F_1 and a lower accuracy at 3°, so the additional time is what buys the accuracy the table reports. MMVP achieves 100% angular accuracy at 3°, 5°, and 10° simultaneously, maintaining this performance through 60% contamination. Only a single decrease to 98.7% at 50% contamination occurs, corresponding to two errors out of 150 directions. Every other error in this range stays below the finest threshold. This pattern changes only at 70% contamination, where the three metrics decrease to 84.7, 88.0, and 93.3 (Table S4).

In contrast, GlobustVP exhibits significant performance degradation as contamination increases, and is no longer competitive beyond 50% outliers. At 60% outliers, it fails completely on 20 out of 50 instances, and its F_1 score of 31.5 is substantially lower than the 87.6 achieved by the proposed method, which is the highest among all methods at this contamination level.

Among the learning-based baselines, CONSAC and PARSAC both fall short of MMVP at the 3° threshold, reaching 90.0 and 24.0, although the two perform similarly on real data. The geometric-prior network [32] and U-ARE-ME are excluded because these approaches require pixel-based inputs, whereas the generator produces only line segments. Rendering line drawings to simulate photographs would evaluate the simulation instead of the underlying method.

Precision and recall across outlier ratios. Figure S5 separates the outlier-robustness study into its two constituent components. This separation aims to eliminate the possibility of a precision and recall trade-off. MMVP demonstrates superior precision across all ratios and maintains higher recall, except at 20% contamination, where it is on a par with Quasi-VP. Consequently, the F_1 ordering does not result from compensation between precision and recall.

G Complete metric tables

G.1 Synthetic results at every outlier ratio

Table S2 presents results for a single operating point (30% outliers). In contrast, Tables S3 and S4 report the same seven metrics across the full range of outlier ratios, from 0% to 70%. MMVP maintains AA@3°/AA@5°/AA@10° at 100.0 from 0% to 60% contamination, with only a minor dip to 98.7 at 50% (two of 150 directions), and first falls below saturation at 70% (84.7/88.0/93.3). Across the 0–60% range, its F_1 score declines smoothly from 95.3 to 87.6. Recall stays at 91.0 or above throughout, indicating that the reduction comes from marginal lines being admitted as inliers, not from errors in direction placement. Competitor methods diverge along this axis. Quasi-VP performs comparably under low contamination but deteriorates as outlier ratios increase, with F_1 decreasing from 95.1 at 0% to 60.5 at 70%. CONSAC follows a similar trend, declining from 92.0 to 60.9. RANSAC demonstrates the greatest durability among competitors, achieving 74.3 at 70%, but consistently underperforms relative to MMVP at all ratios. PARSAC drops to 50.2 already at 20%. Linkage methods fail between 30% and 40% (F_1 46.2 → 26.0 for T-Linkage), while the relaxation-based pipeline fails between 50% and 60%, with outright failure in 20 of 50 instances at 60%, increasing further at 70%. Consequently, the performance margin over the nearest competitor widens as contamination increases, from less than one point at 30% to 9.4 at 60%.

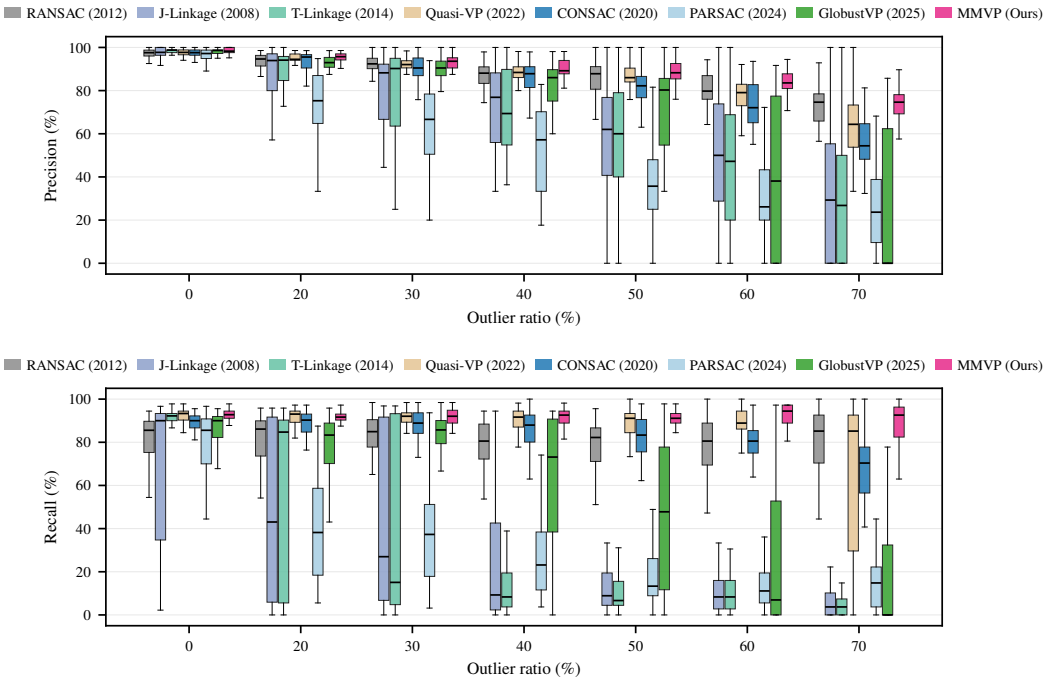

Figure S5: **Precision and recall behind the F_1 trend** (synthetic, $m = 90$, $\sigma = 1$ px, 50 trials), for all eight methods evaluable on line input. Separating the two shows that MMVP's F_1 lead is not a precision/recall trade, since it leads on precision throughout and on recall everywhere save at 20%, where it is on a par with the closest competitor, and its decline past 50% contamination is driven by precision, as the method admits marginal lines.

Table S3: **Synthetic scenes, outlier ratios** 0–40%. $m = 90$ lines, $\sigma = 1$ px endpoint noise, 50 seeded trials per setting. Best per column in **bold**, second best underlined.

Method	Prec. ↑	Rec. ↑	F_1 ↑	AA@3° ↑	AA@5° ↑	AA@10° ↑	time ↓
outlier ratio $= 0\%$							
J-Linkage (2008) [14]	95.9	70.7	76.5	78.0	79.3	79.3	0.003
RANSAC (2012) [12]	97.4	82.3	88.9	86.7	93.3	97.3	**0.000**
T-Linkage (2014) [15]	97.5	90.2	93.3	98.0	98.0	98.0	0.003
CONSAC (2020) [29]	96.2	88.2	92.0	96.0	96.7	97.3	11.652
Quasi-VP (2022) [17]	97.9	92.5	95.1	**100.0**	**100.0**	**100.0**	0.001
PARSAC (2024) [30]	95.4	79.7	86.2	79.3	82.0	84.7	0.010
GlobustVP, MOSEK (2025) [19]	96.9	85.5	90.4	93.3	95.3	98.7	0.345
MMVP (Ours)	**98.1**	**92.6**	**95.3**	**100.0**	**100.0**	**100.0**	0.029
outlier ratio $= 20\%$							
J-Linkage (2008) [14]	85.9	49.8	54.3	52.0	52.7	58.7	0.003
RANSAC (2012) [12]	93.7	81.7	87.0	86.0	96.0	97.3	0.001
T-Linkage (2014) [15]	84.9	53.9	57.5	54.7	59.3	64.0	0.003
CONSAC (2020) [29]	92.2	86.7	89.3	86.0	88.7	90.7	11.049
Quasi-VP (2022) [17]	95.1	**91.9**	**93.5**	98.0	**100.0**	**100.0**	**0.001**
PARSAC (2024) [30]	72.4	40.6	50.2	30.0	33.3	42.0	0.010
GlobustVP, MOSEK (2025) [19]	91.9	75.9	81.9	76.0	84.7	90.0	0.822
MMVP (Ours)	**95.4**	91.4	93.3	**100.0**	**100.0**	**100.0**	0.034
outlier ratio $= 30\%$							
J-Linkage (2008) [14]	77.1	46.2	50.7	46.0	49.3	54.7	0.003
RANSAC (2012) [12]	92.5	82.8	87.1	89.3	94.0	97.3	0.001
T-Linkage (2014) [15]	76.5	43.0	46.2	42.0	46.7	50.7	0.003
CONSAC (2020) [29]	89.4	86.7	87.9	90.0	90.7	92.0	10.814
Quasi-VP (2022) [17]	92.1	91.4	91.7	98.7	**100.0**	**100.0**	**0.001**
PARSAC (2024) [30]	64.0	36.4	44.9	24.0	30.0	38.0	0.010
GlobustVP, MOSEK (2025) [19]	89.5	77.8	81.8	81.3	85.3	90.0	1.590
MMVP (Ours)	**93.3**	**91.7**	**92.5**	**100.0**	**100.0**	**100.0**	0.034
outlier ratio $= 40\%$							
J-Linkage (2008) [14]	64.6	25.3	31.0	21.3	27.3	33.3	0.003
RANSAC (2012) [12]	87.0	79.1	82.6	77.3	92.0	97.3	0.002
T-Linkage (2014) [15]	67.6	20.8	26.0	16.0	24.0	30.7	0.003
CONSAC (2020) [29]	85.6	85.8	85.6	80.7	84.7	86.0	10.769
Quasi-VP (2022) [17]	88.9	90.2	89.4	94.7	98.7	**100.0**	**0.000**
PARSAC (2024) [30]	52.6	27.9	35.3	10.0	15.3	23.3	0.010
GlobustVP, MOSEK (2025) [19]	76.6	63.0	67.1	62.0	64.7	69.3	2.973
MMVP (Ours)	**90.1**	**91.3**	**90.6**	**100.0**	**100.0**	**100.0**	0.033

Table S4: **Synthetic scenes, outlier ratios** 50–70%. $m = 90$ lines, $\sigma = 1$ px endpoint noise, 50 seeded trials per setting. Best per column in **bold**, second best underlined.

Method	Prec. ↑	Rec. ↑	F_1 ↑	AA@3° ↑	AA@5° ↑	AA@10° ↑	time ↓
outlier ratio $= 50\%$							
J-Linkage (2008) [14]	55.3	18.4	23.6	13.3	19.3	26.7	0.003
RANSAC (2012) [12]	86.0	79.9	82.6	77.3	90.7	93.3	0.002
T-Linkage (2014) [15]	58.2	15.1	20.5	8.7	14.7	20.7	0.003
CONSAC (2020) [29]	82.0	82.3	82.0	71.3	76.0	78.0	10.692
Quasi-VP (2022) [17]	86.0	87.8	86.6	90.7	92.7	96.0	**0.001**
PARSAC (2024) [30]	37.9	19.0	24.6	8.0	9.3	17.3	0.011
GlobustVP, MOSEK (2025) [19]	63.0	47.2	52.0	37.3	48.0	54.7	4.663
MMVP (Ours)	**88.5**	**91.0**	**89.6**	**98.7**	**98.7**	**98.7**	0.034
outlier ratio $= 60\%$							
J-Linkage (2008) [14]	47.8	13.7	19.1	7.3	12.0	16.7	0.003
RANSAC (2012) [12]	79.8	76.4	77.4	67.3	84.7	91.3	0.004
T-Linkage (2014) [15]	44.7	12.4	17.6	2.0	5.3	12.7	0.004
CONSAC (2020) [29]	72.9	79.7	76.0	62.0	70.0	74.0	10.696
Quasi-VP (2022) [17]	76.5	82.0	78.2	80.0	87.3	88.0	**0.000**
PARSAC (2024) [30]	30.6	16.2	20.6	6.0	8.0	12.7	0.011
GlobustVP, MOSEK (2025) [19]	37.7	29.0	31.5	27.3	29.3	34.0	6.892
MMVP (Ours)	**83.8**	**91.9**	**87.6**	**100.0**	**100.0**	**100.0**	0.033
outlier ratio $= 70\%$							
J-Linkage (2008) [14]	33.9	7.8	11.7	1.3	4.0	8.0	0.003
RANSAC (2012) [12]	72.2	78.0	74.3	78.7	85.3	90.0	0.006
T-Linkage (2014) [15]	28.0	4.9	8.1	0.0	0.0	3.3	0.003
CONSAC (2020) [29]	56.1	67.1	60.9	52.7	56.0	60.7	10.706
Quasi-VP (2022) [17]	60.5	64.1	60.5	58.0	64.0	70.7	**0.000**
PARSAC (2024) [30]	25.5	17.9	20.6	5.3	7.3	12.7	0.011
GlobustVP, MOSEK (2025) [19]	25.4	18.8	20.6	12.7	17.3	20.0	10.668
MMVP (Ours)	**73.9**	**84.1**	**77.6**	**84.7**	**88.0**	**93.3**	0.032

Table S5: YUD+ [29], all 102 images. Best per column in **bold**, second best underlined.

Method	Prec. ↑	Rec. ↑	F_1 ↑	AA@3° ↑	AA@5° ↑	AA@10° ↑	time ↓
J-Linkage (2008) [14]	71.4	34.6	41.8	22.6	27.7	33.6	0.255
RANSAC (2012) [12]	87.3	79.5	83.0	54.0	69.8	83.6	**0.001**
T-Linkage (2014) [15]	61.2	28.3	34.3	14.4	23.2	33.1	0.258
CONSAC (2020) [29]	92.4	93.2	92.8	78.2	81.9	85.6	11.792
Quasi-VP (2022) [17]	90.8	90.6	90.6	73.2	82.2	86.4	0.010
Lin et al. (2022) [32]	90.9	83.6	86.8	65.0	70.1	76.3	0.123
PARSAC (2024) [30]	94.1	94.0	94.0	78.2	83.3	86.2	0.011
U-ARE-ME (2024) [33]	86.4	74.5	79.4	57.3	79.9	87.6	0.035
GlobustVP, MOSEK (2025) [19]	63.0	52.4	56.3	32.5	44.4	53.1	1.255
MMVP (Ours)	**95.2**	**96.0**	**95.5**	**83.3**	**86.4**	**88.1**	0.066

H Real data: YUD+, where the Manhattan assumption breaks

All results presented thus far assume exactly three orthogonal directions, which is the assumption encoded by the objective. YUD+ [29] enables measurement of the impact when this assumption does not hold, as it re-annotates the 102 calibrated urban images of the York Urban dataset with additional vanishing points. Of these, 71 images contain three directions, while 31 images contain between four and eight, with both subsets captured under identical camera and environmental conditions. Since each method evaluated returns exactly three orthogonal directions, angular scoring is recall-oriented, in that for each ground-truth vanishing point the angle to the closest estimated direction is measured, following [29; 30]. Association metrics are computed with respect to the three dominant ground-truth directions, which is exact for the 71 Manhattan images and does not account for the additional directions in the remaining 31 images.

As shown in Table S5, MMVP achieves the highest precision, recall, F_1 score, and performance at all three angular thresholds, consistent with results on both synthetic data and HoliCity. MMVP surpasses PARSAC, the strongest learning-based competitor on this dataset, by 1.1 points in precision, 2.0 in recall and 5.1 at the 3° threshold, without any training. The wall-clock time reported for PARSAC reflects a GPU-batched network, whereas our results are obtained using a CPU-based solver. In the accuracy columns, where hardware acceleration does not confer an advantage, MMVP consistently outperforms the alternatives. The performance disparity between the two real datasets is also informative. PARSAC's score at the 3° threshold drops from 78.2 on YUD+ to 54.6 on HoliCity, and CONSAC's from 78.2 to 58.3, while MMVP decreases only from 83.3 to 68.7. This stability demonstrates that an estimator requiring no training incurs a minimal domain-transfer penalty.

Explanation for the larger margin compared to curated line sets. YUD+ is evaluated on approximately 550 detected segments per image, resulting in a substantial proportion of outliers or noisily terminated fragments. This scenario aligns with the conditions described in the noise study of Appendix F, which predicts that a non-minimal estimator will be preferred. This prediction is confirmed by the observed ordering, since a three-line minimal hypothesis inherits the complete endpoint error from its three segments, whereas the weighted rotation step described in Section 3 distributes this error across all lines permitted by the association. Consequently, the advantage increases as input contamination rises, a condition typical of the output of real line detectors.

Figure S6 presents the corresponding gallery for YUD+. In the middle row, the two linkage methods achieve almost no correct associations, with F_1 scores of 0.0 and 0.2. The bottom row depicts a non-Manhattan scene containing four ground-truth directions. Here the largest angular error is 37.8°, while the association reaches an F_1 of 97.2. This high score occurs because the missed element is the fourth direction, which cannot be represented by any three-direction solver, not a poorly estimated rotation. All methods in the comparison yield angular errors between 36.9 and 38.6° in this scenario. Additional panels for both datasets are provided in Appendix I.

Table S6 divides YUD+ into two subsets, 71 images annotated with exactly three vanishing points and 31 images annotated with four to eight vanishing points. This division highlights the dataset's purpose, as a solver restricted to three orthogonal directions cannot represent a fourth. Consequently, the second subset quantifies the cost of a violated Manhattan assumption for every three-direction



Figure S6: Qualitative results on YUD+, one row per image. Colouring and annotation as in Figure 2.

method, not the effectiveness of any one of them. MMVP keeps its association accuracy across this division, with an F_1 of 95.5 on the Manhattan images and 95.7 on the others, and leads at every angular threshold in both subsets.

I Additional qualitative panels

Figures S7–S10 for YUD+ and Figures S11–S14 for HoliCity present twelve scenes per dataset. Each comparison method is illustrated on the same detected segments, with its association F_1 and its worst-direction angular error indicated above each panel.

These figures reveal two aspects that an average over a test split cannot convey. First, they illustrate the extent to which the methods diverge on individual images. MMVP achieves an association F_1 between 83 and 99 across these scenes. In contrast, the weakest competitor fails to return results on nearly half of the same scenes, with the relaxation-based solver unable to produce a frame and the linkage methods assigning almost no lines to any direction. Thus, a small difference between two column averages can represent the distinction between a method that consistently provides results and one that does so intermittently. Second, the association and angular error metrics address different questions and may diverge. For example, in the non-Manhattan scenes of Figure S10 and the most challenging crops of Figure S14, the association remains above 90, while the worst-direction error is several tens of degrees. The panels attribute this discrepancy to the scene itself, not to the estimation,

Table S6: **YUD+ split by whether the Manhattan assumption holds.** The same seven metrics as Table S5, separated into the 71 images with exactly three ground-truth VPs and the 31 with four to eight. Best per column in **bold**, second best underlined.

Method	Prec. ↑	Rec. ↑	F_1 ↑	AA@3° ↑	AA@5° ↑	AA@10° ↑	time ↓
all images, $n = 102$ images							
J-Linkage (2008) [14]	71.4	34.6	41.8	22.6	27.7	33.6	0.255
RANSAC (2012) [12]	87.3	79.5	83.0	54.0	69.8	83.6	**0.001**
T-Linkage (2014) [15]	61.2	28.3	34.3	14.4	23.2	33.1	0.258
CONSAC (2020) [29]	92.4	93.2	92.8	78.2	81.9	85.6	11.792
Quasi-VP (2022) [17]	90.8	90.6	90.6	73.2	82.2	86.4	0.010
Lin et al. (2022) [32]	90.9	83.6	86.8	65.0	70.1	76.3	0.123
PARSAC (2024) [30]	94.1	94.0	94.0	78.2	83.3	86.2	0.011
U-ARE-ME (2024) [33]	86.4	74.5	79.4	57.3	79.9	87.6	0.035
GlobustVP, MOSEK (2025) [19]	63.0	52.4	56.3	32.5	44.4	53.1	1.255
MMVP (Ours)	**95.2**	**96.0**	**95.5**	**83.3**	**86.4**	**88.1**	0.066
Manhattan holds (exactly 3 GT VPs), $n = 71$ images							
J-Linkage (2008) [14]	72.6	36.0	42.7	28.6	34.3	40.8	0.232
RANSAC (2012) [12]	87.4	79.9	83.3	61.0	80.3	94.4	**0.001**
T-Linkage (2014) [15]	61.8	28.5	34.6	17.8	26.8	38.5	0.227
CONSAC (2020) [29]	93.0	93.9	93.4	88.7	93.4	96.7	11.743
Quasi-VP (2022) [17]	90.9	91.4	91.1	85.9	94.8	99.1	0.010
Lin et al. (2022) [32]	90.8	83.1	86.5	72.3	78.9	85.0	0.123
PARSAC (2024) [30]	94.1	94.3	94.2	89.7	94.8	97.7	0.011
U-ARE-ME (2024) [33]	85.9	74.7	79.3	66.7	92.0	99.5	0.035
GlobustVP, MOSEK (2025) [19]	65.0	54.6	58.6	40.8	54.9	62.0	1.252
MMVP (Ours)	**94.9**	**96.1**	**95.5**	**95.3**	**99.1**	**100.0**	0.064
Manhattan violated (4–8 GT VPs), $n = 31$ images							
J-Linkage (2008) [14]	68.7	31.3	39.7	13.5	17.7	22.7	0.386
RANSAC (2012) [12]	87.0	78.7	82.4	43.3	53.9	67.4	**0.001**
T-Linkage (2014) [15]	60.0	27.8	33.8	9.2	17.7	24.8	0.369
CONSAC (2020) [29]	91.0	91.7	91.3	62.4	64.5	68.8	11.852
Quasi-VP (2022) [17]	90.4	88.7	89.4	53.9	63.1	67.4	0.014
Lin et al. (2022) [32]	91.3	84.7	87.7	53.9	56.7	63.1	0.123
PARSAC (2024) [30]	94.0	93.4	93.7	61.0	66.0	68.8	0.011
U-ARE-ME (2024) [33]	87.4	74.1	79.7	43.3	61.7	69.5	0.035
GlobustVP, MOSEK (2025) [19]	58.3	47.3	51.2	19.9	28.4	39.7	1.405
MMVP (Ours)	**95.7**	**95.7**	**95.7**	**65.2**	**67.4**	**70.2**	0.070

as the missed direction is either a fourth one, which cannot be represented by a three-direction model, or one supported by only a few detected segments, insufficient to constrain it or to influence a score computed over all segments. All three-direction methods in the comparison encounter this limitation, which is why Table S6 reports the two regimes separately.

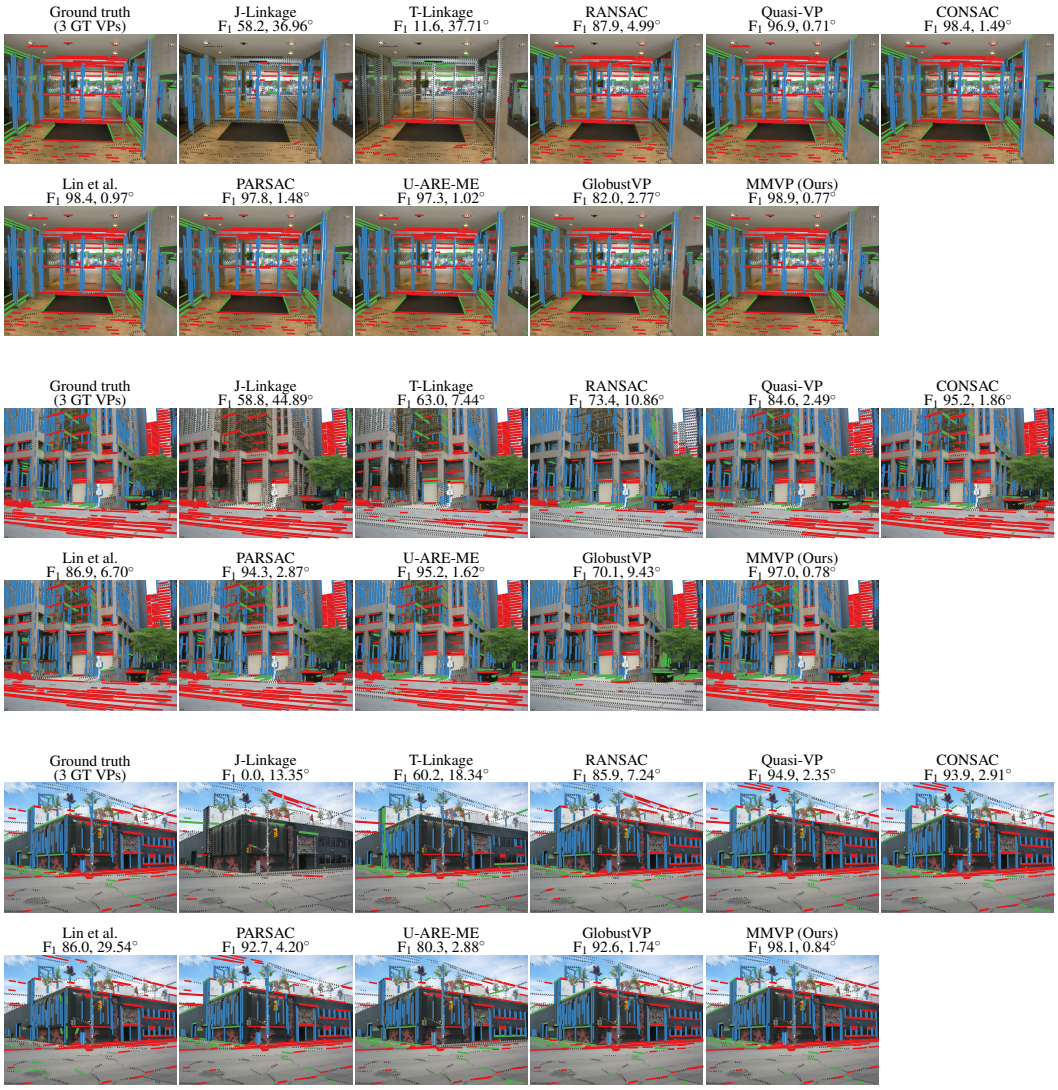

Figure S7: Further qualitative results on YUD+. Lines are coloured by the recovered line–VP association, dotted black for outliers. All three are Manhattan scenes, with exactly three ground-truth vanishing points, recovered to under 1°.

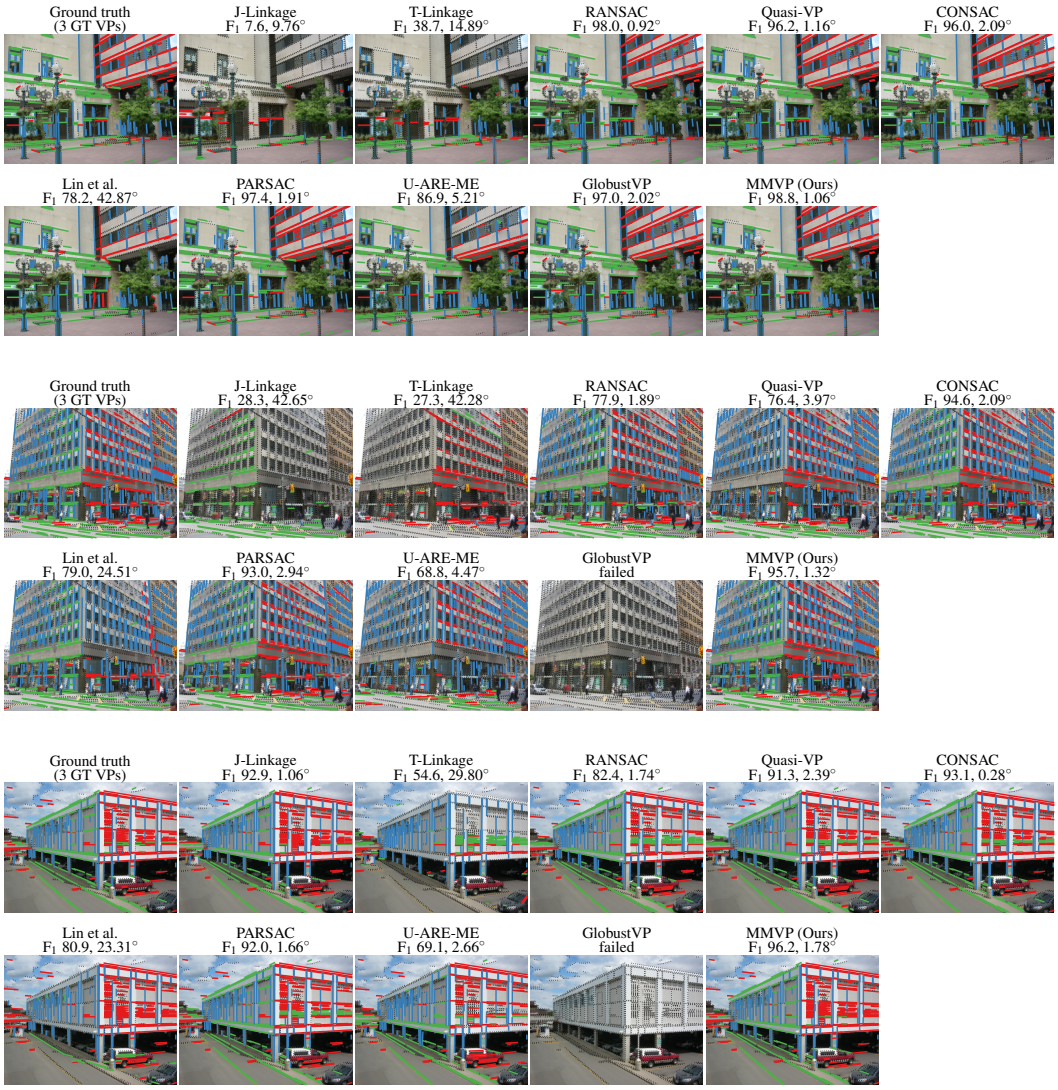

Figure S8: Further qualitative results on YUD+. Colouring and annotation as in Figure S7. Manhattan scenes recovered to 1.1–1.8°.

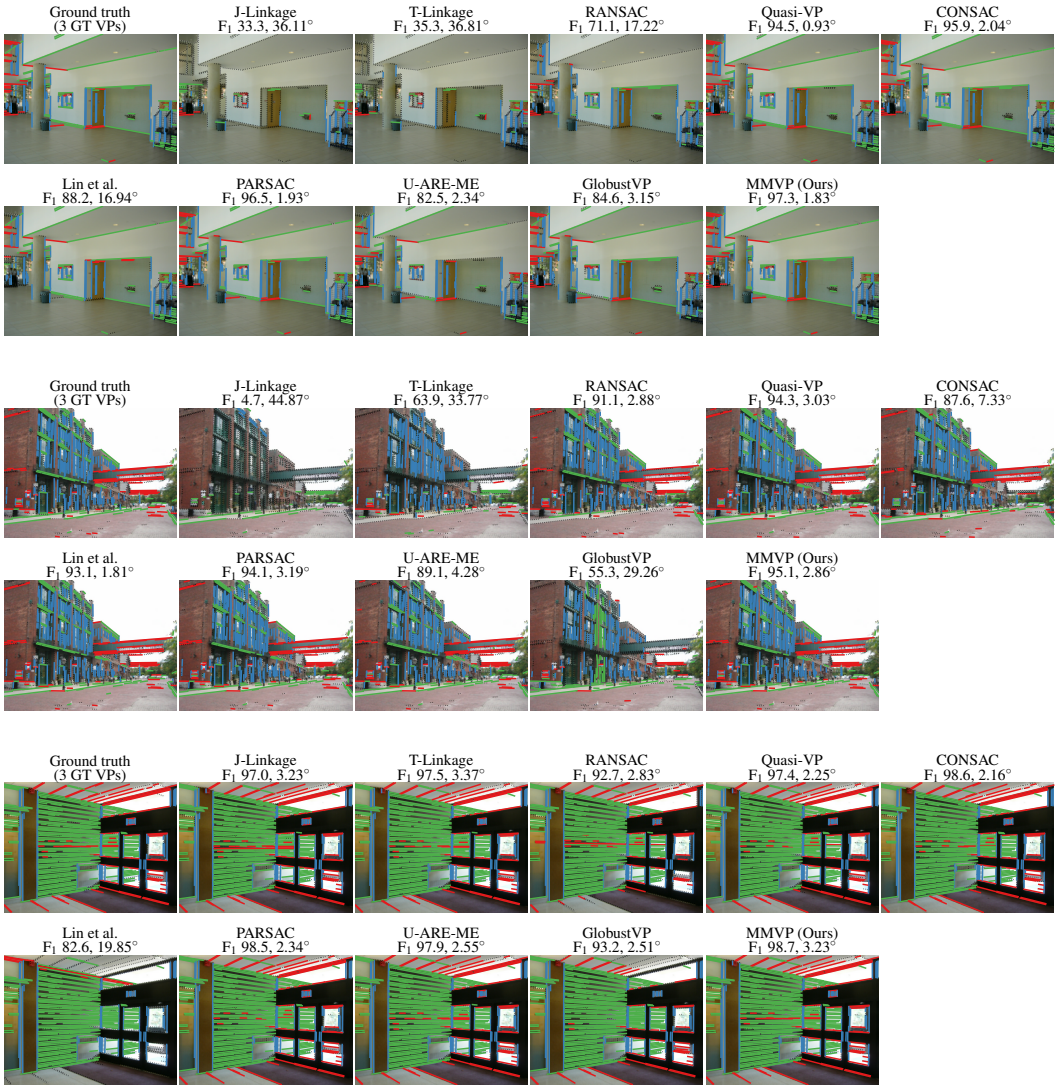

Figure S9: Further qualitative results on YUD+. Colouring and annotation as in Figure S7. Manhattan scenes recovered to 1.8–3.2°.

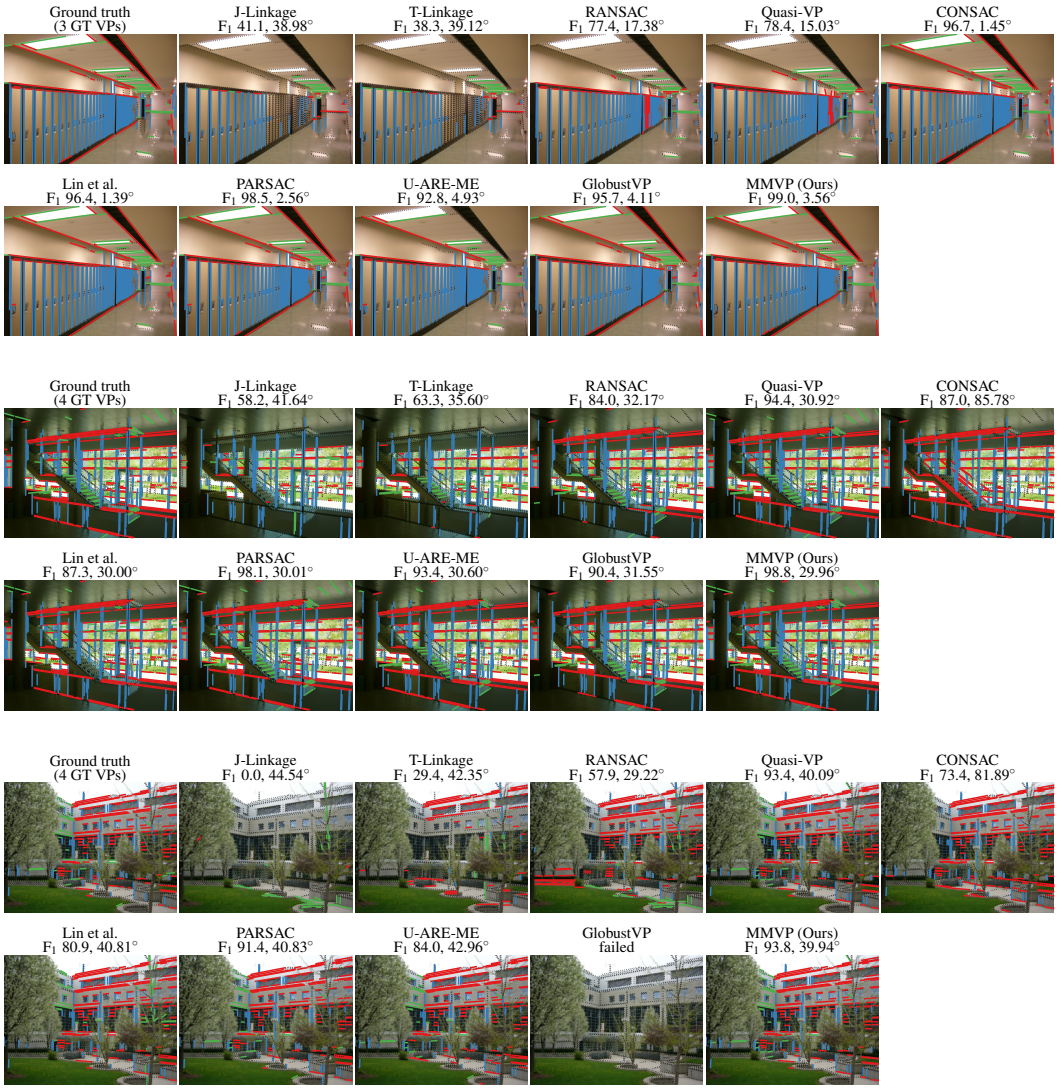

Figure S10: Further qualitative results on YUD+. Colouring and annotation as in Figure S7. The second and third scenes are non-Manhattan, their annotation carrying *four* vanishing points instead of three, and MMVP attains the best association F_1 on each (98.8 and 93.8). Table S6 quantifies the effect over all 31 such images.



Figure S11: Further qualitative results on HoliCity. Colouring and annotation as in Figure S7. All three crops are recovered to about a degree (0.6–1.1°).

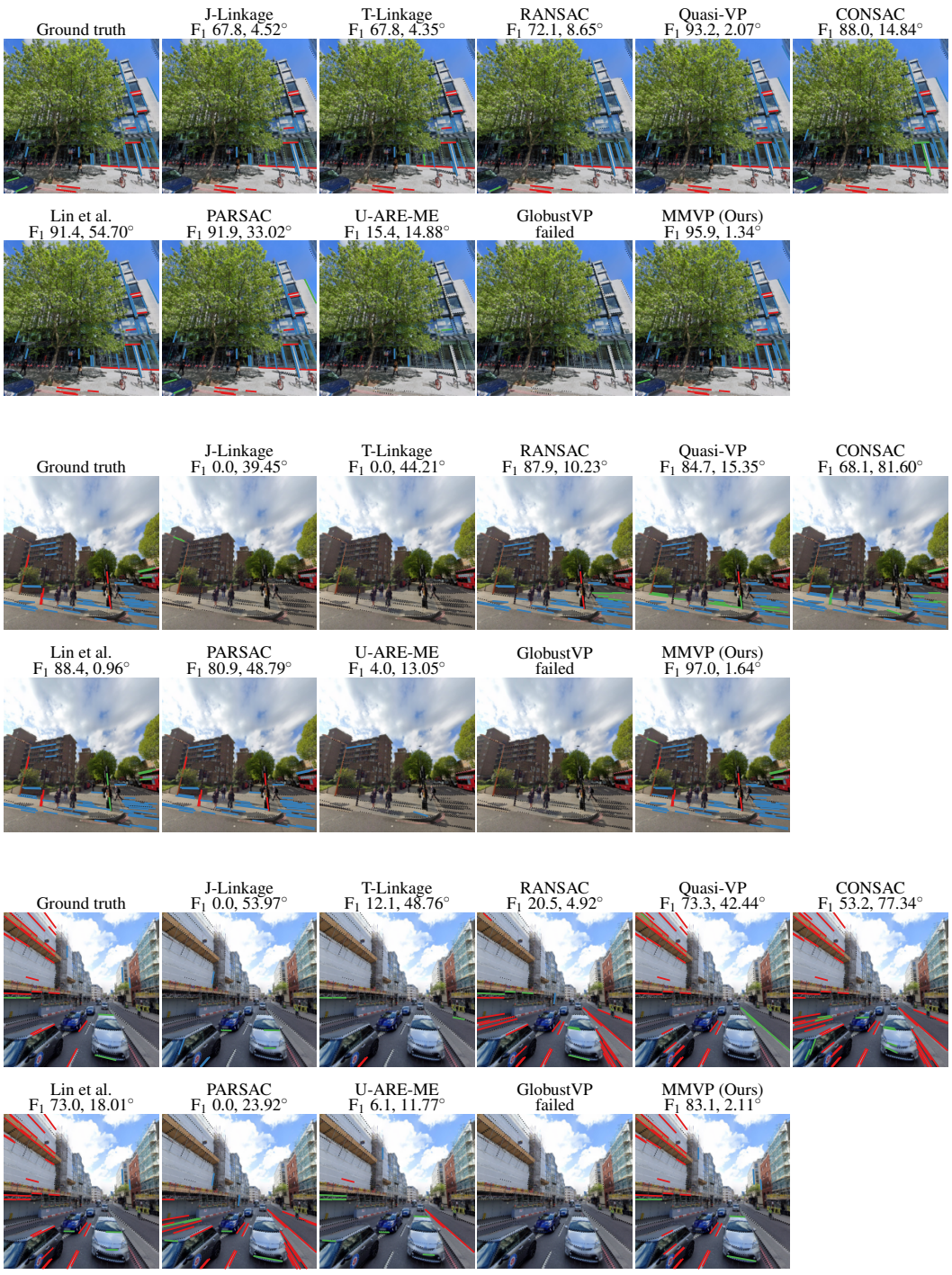

Figure S12: Further qualitative results on HoliCity. Colouring and annotation as in Figure S7. Worst-direction error 1.3–2.1°.



Figure S13: Further qualitative results on HoliCity. Colouring and annotation as in Figure S7. Worst-direction error 2.9–11.1°, with MMVP still leading on association F_1.

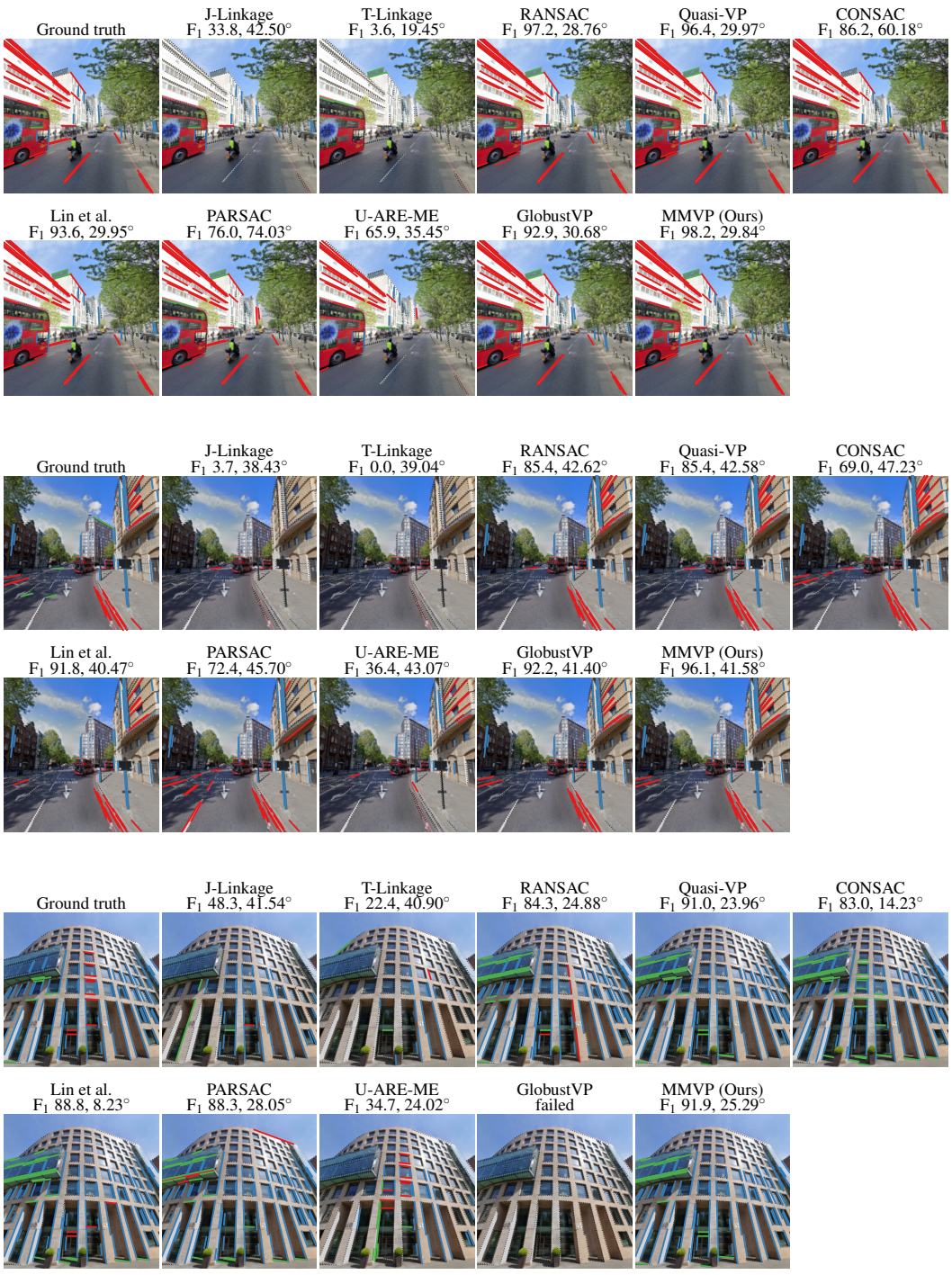

Figure S14: Further qualitative results on HoliCity. Colouring and annotation as in Figure S7. These are the clearest cases for reading the two annotations together instead of either alone, since MMVP holds the best association F_1 (91.9–98.2) while its worst direction is 25–42° off.