Robust Incremental Structure-from-Motion with Hybrid Features

Shaohui Liu[1⋆], Yidan Gao[1*], Tianyi Zhang[1*], Rémi Pautrat[2], Johannes L. Schönberger[2], Viktor Larsson[3], and Marc Pollefeys[1,2]

[1]ETH Zurich [2]Microsoft Mixed Reality & AI Lab Zurich [3]Lund University

Abstract. Structure-from-Motion (SfM) has become a ubiquitous tool for camera calibration and scene reconstruction with many downstream applications in computer vision and beyond. While the state-of-the-art SfM pipelines have reached a high level of maturity in well-textured and well-configured scenes over the last decades, they still fall short of robustly solving the SfM problem in challenging scenarios. In particular, weakly textured scenes and poorly constrained configurations oftentimes cause catastrophic failures or large errors for the primarily keypoint-based pipelines. In these scenarios, line segments are often abundant and can offer complementary geometric constraints. Their large spatial extent and typically structured configurations lead to stronger geometric constraints as compared to traditional keypoint-based methods. In this work, we introduce an incremental SfM system that, in addition to points, leverages lines and their structured geometric relations. Our technical contributions span the entire pipeline (mapping, triangulation, registration) and we integrate these into a comprehensive end-to-end SfM system that we share as an open-source software with the community. We also present the first analytical method to propagate uncertainties for 3D optimized lines via sensitivity analysis. Experiments show that our system is consistently more robust and accurate compared to the widely used point-based state of the art in SfM – achieving richer maps and more precise camera registrations, especially under challenging conditions. In addition, our uncertainty-aware localization module alone is able to consistently improve over the state of the art under both point-alone and hybrid setups.

1 Introduction

Estimating camera parameters and scene geometry from images, also known as Structure-from-Motion (SfM), has enabled a wide variety of applications such as augmented reality [77, 80], novel view synthesis [41, 59], scene reconstruction [48,107], *etc*. For SfM, the incremental paradigm that alternates between updating the map and resectioning cameras is by far the most popular. This is due to its comparably better accuracy and robustness, as well as having an active open-source community with multiple well-engineered pipelines [79, 86, 102], of which COLMAP [79] has become the de-facto standard SfM in the recent years.

⋆ Equal contribution

Fig. 1: Incremental Structure-from-Motion with points, lines, and vanishing points. Red: groundtruths. **Blue:** predictions. We show example indoor scenes where classical point-based Structure-from-Motion fails. Leveraging additional constraints from line segments, our pipeline can faithfully reconstruct the scene and cameras.

Most modern SfM systems heavily rely on the presence of stable feature points in the scene, which are detected, matched, and triangulated into sparse 3D point cloud maps followed by repeated bundle adjustments and camera pose estimations. This, however, regularly prevents these systems from providing robust and accurate results in poorly conditioned scenarios, where the scene has little texture and thus few feature points (*e.g.*, indoor scenes). Compared to points, lines frequently appear in human-made environments where feature points are sparse, they have a large spatial extent, and they often appear in structured configurations offering additional geometric constraints (parallelism, orthogonality, *etc.*). The idea to exploit more structured features, such as straight lines, dates back to the early 2000s [6, 7, 14, 78, 93].

Despite these clear advantages, line segments are not used in the currently available state-of-the-art SfM pipelines. This is mainly due to line reconstruction coming with additional challenges compared to the point-based counterparts. For example, lines are in general more difficult to describe due to often having inconsistent endpoints across views. In practice, lines also suffer from unstable degenerate configurations during triangulation (as extensively studied in [52]). Moreover, while there has been great progress in feature point detection and matching in the past decade, line detection and matching has received comparably less attention. However, recently, significant progress has been made on line detectors [34, 65, 68, 105] and matchers [1, 67] thanks to the advent of deep learning, making it possible to revisit lines as features in SfM.

In this work, we introduce an end-to-end incremental SfM system that consistently improves the robustness and accuracy over the state of the art by incorporating hybrid features, including points, lines, vanishing points (VPs), and their structured relations. Our technical contributions span across all the three main steps of incremental SfM: triangulation, refinement, and registration, introducing new robust mechanisms to reliably maintain structural features, and incorporating uncertainty measurements to further improve the robustness. Our system is consistently more robust and accurate compared to the widely used SfM pipeline COLMAP [79], achieving more precise camera localization, richer sparse maps, more valid registrations, and less catastrophic failure cases (*cf*. Fig. 1). By

sharing our code as an open-source software, we hope to enable further research on SfM as well as benefit downstream applications in the community.

Specifically, our technical contributions are listed as follows:

- **System:** We present the first end-to-end incremental SfM system that rigorously integrates points, lines, vanishing points (VPs), and their relations.
- **Incremental Mapping:** We extend incremental triangulation operations initially designed for point features to lines and VPs, leading to an incremental line triangulator with comparable performance to the global triangulator in [52]. This removes the need from [52] to get all the images posed beforehand.
- **Refinement and Bundle Adjustment:** We propose to explicitly identify reliable/unreliable tracks with uncertainty modeling and apply two-step refinement with cached inactive supports. This prevents prematurely filtering unreliable line tracks in early stages, without sacrificing the pose accuracy.
- **Registration with Hybrid Features:** We estimate the 6-DoF camera pose with points, lines and vanishing points together in a hybrid RANSAC framework, employing three existing point-line solvers [113] and two extra gravity-based solvers from a VP correspondence.
- **Uncertainty Modeling for 3D Maps:** We perform uncertainty propagation for both 3D points and lines. In particular, this paper introduces the first analytical method to propagate uncertainties for 3D optimized lines.
- **Uncertainty Integration for Refinement and Registration:** We successfully integrate uncertainty in both refinement and registration. Our uncertainty-based registration improves upon the state of the art on public localization benchmarks under both point-alone [75] and hybrid [52] cases.

2 Related Work

Structure from Motion. Incremental methods have traditionally dominated the state of the art in SfM in terms of robustness and accuracy with an active research community and several open-source software packages [60, 79, 86, 102]. Different from global SfM [16, 36, 60, 85, 90, 91, 101], incremental methods sequentially register images followed by repeated local and global refinements. This approach is usually slower but yields more robust and accurate results. The community has made tremendous progress on efficiency and scalability [2, 9, 16, 24, 44, 87, 103] as well as robustness and accuracy [15, 17, 19, 20, 46, 63, 72, 79]. Over the last years, COLMAP [79] has emerged as the de facto standard incremental pipeline for SfM, with applications in many downstream tasks in computer vision [41, 59] and beyond. Most recently, and orthogonal to our contributions, learning-based pipelines have also been explored [92, 97–99, 104, 110], yet still being unable to match the performance of COLMAP on large-scale scenes. Improvements on using pixel-perfect features [51] and semi-dense matching [30] show great potential and could be combined with our work. In this paper, we introduce a scalable incremental SfM system, that is built upon the success of COLMAP, while improving its robustness and accuracy by carefully incorporating structural features into the entire reconstruction process.

Integration of Lines/Structures in Geometric Pipelines. The idea of improving SfM with line features dates back to the early 2000s. Bartoli and Sturm [7] pioneered a full SfM system with lines, followed by Schindler [78] who integrated Manhattan assumptions. Chandraker *et al*. [14] proposed a robust stereo-based system on infinite lines. Work in the field of SLAM and visual odometry [27, 28, 31, 49, 50, 55, 56, 71, 84, 100, 106, 114] has focused on integrating line features to improve the accuracy, yet only constrained to sequential data and often with strong motion assumption (*e.g.*, from inertial data), while our approach works on general, unstructured input. In recent years, researchers have also made progress on incorporating lines and vanishing points (VP) in global methods [33, 58] and exploiting curves in bundle adjustment [64]. However, none of the previous works have developed a full end-to-end SfM system that can in practice compete with COLMAP [79] in terms of versatility and robustness. While lines intuitively provide benefits in terms of complementary and rich geometric constraints [8, 23, 47, 52, 66, 73, 111], they come with significant practical challenges [14, 52] due to occlusion of endpoints and degenerate configurations as well as inherent difficulties to match them robustly across different views. The recent breakthrough developments on line detectors [65, 68, 96, 105] and matchers [1, 65, 67], spur a renewed interest in the community to revisit the problem of leveraging lines and their structural configurations [32, 52], which is further approached with a learning-based solution in [11]. These prior works have focused on 3D line reconstruction from given camera geometry. In contrast, we develop a general method that solves the full SfM problem and thus jointly benefits the robustness and accuracy of SfM estimation.

Uncertainties in Multi-View Geometry. Modeling of uncertainty is a long-standing and important problem in computer vision [5, 10, 22]. Throughout the years, researchers have continuously made progress on modeling the uncertainty of local point features [18,38,39,57,88,108] and their matches [26,61,109]. There were also attempts on incorporating uncertainty measurements for radar odometry [12], 3D benchmark construction [77], multi-view stereo [42, 70, 82, 112], *etc*. Despite the breadth and depth of research in this domain, uncertainty modeling remains challenging in practice and has not been embedded in a principled manner in most SfM pipelines. Our work integrates principled uncertainty modeling into the reconstruction process for improved robustness and accuracy. This is important for the integration of lines, which, as we show, especially benefit from probabilistic modeling. Building on top of the Jacobian derivation on the line reprojection error [7, 114], we propose the first method to analytically model the uncertainty of 3D optimized lines based on sensitivity analysis [21].

3 Methodology

In this section, we present our proposed SfM pipeline. Our method takes an unordered image collection as input, with either calibrated or uncalibrated camera

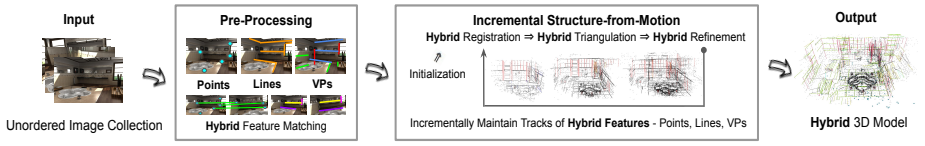


Fig. 2: Our proposed SfM pipeline exploits hybrid features including points, lines, and vanishing points (VPs). We improve with technical contributions over all three main components: registration, triangulation, and refinement, leading to richer 3D maps (with uncertainty measurements) and more robust camera localization.

intrinsics. To be able to detect and match straight lines, we require the images to have no radial or tangential distortion.

Fig. 2 shows an overview of the proposed pipeline. Our setup and overall system design leverages the same incremental reconstruction approach as the popular point-based SfM software COLMAP [79]. The system first performs local feature detection on the input images and then matches pairs of images with different strategies (exhaustive, sequential, *etc.*). Next, the procedure bootstraps the reconstruction process with an initial image pair, followed by progressively adding new images by alternating between camera registration, triangulation of newly observed structures, and iterative local and global refinement using bundle adjustment. The output of our system is a set of estimated camera parameters and a sparse 3D map with hybrid local features: points, lines, VPs, and their geometric relations. In the following parts, we detail the design of the three main modules: mapping (Sec. 3.1), refinement (Sec. 3.2), and registration (Sec. 3.3).

Detection and Matching on Images. Instead of solely relying on point features as in [79, 86, 102], we additionally detect and match lines and vanishing points (VPs) across images. The detection and matching of lines can benefit from any existing detectors and matchers. We take the state-of-the-art DeepLSD [68] detector and GlueStick [67] matcher as our default choices. For VPs, we use JLinkage [94] as the default detector. The two-view matching of VPs is done through consensus voting from the matches of their associated lines. We consider two VPs a good match if they share at least five line matches.

3.1 Incremental Update of Hybrid Maps

While incremental triangulation of points has reached a high level of maturity throughout the years [29, 54, 79], the triangulation of lines and VPs has not been studied in great detail. This is partially due to the natural challenges of line triangulation, including inconsistent endpoints across views and occlusions as well as more frequent unstable and degenerate view configurations (Fig. 3 shows some examples). While LIMAP [52] recently introduced a global line mapper that can robustly construct line maps from pre-computed posed images, incremental triangulation of lines is drastically harder and requires a sophisticated update mechanism. Different from points, the verification of a 3D line triangulation requires at least three views, making it much more unstable when only a few views are available. This happens especially in the early stage of the incremental triangulation process or in scenes with sparse view coverage.

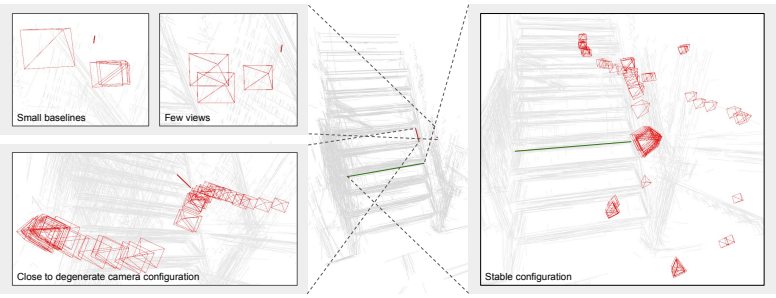


Fig. 3: Line triangulation is sensitive to view configurations. **Left:** unstable tracks from small baselines, few supports, or degenerate patterns. **Right:** an example stable track.

To address these challenges, we combine the techniques from global line triangulation in LIMAP [52] and classic point triangulation [79] to build a robust and efficient hybrid incremental triangulator for points, lines, and VPs.

Triangulating Lines from a New Image. Since line triangulation is inherently unstable and cannot be verified by only two views, we require at least two additional views to triangulate a 3D line. Thus, the incremental triangulation of lines only starts after we register the fourth image. Equivalent to incremental point triangulation, when a new image is registered, we aim to grow the currently triangulated line map with the 2D line features from the new image using the following two operations:

- **Continue**: *extends an existing line track.* Given a new 2D line detection l, we first test if there exists a matched line (in the previously registered images) that is already triangulated in the map and test if the reprojection $\pi(L)$ of the corresponding 3D line L agrees with l. Since there can be several such 3D lines, we first compute a score for each of the candidate 3D lines L in a similar fashion as [52]. Denoting d_{perp} the 2D perpendicular distance and d_{ang} the 2D angular distance defined in [52], and τ_p and τ_a two thresholds (2 pixels and 5 degrees by default), our scoring function is as follows:

$$s(L, l) = \min(e^{-(d_{perp}(\pi(L),l)/\tau_p)}, e^{-(d_{ang}(\pi(L),l)/\tau_a)}). \tag{1}$$

  Similar to [52] we additionally check if there is a sufficient overlap. We select the 3D line L with the highest score, and if both errors are below the thresholds, we add the line l into the corresponding line track.
- **Create**: *triangulates a new line.* If we are unable to assign a line segment to any existing track, we try to create a new 3D line with two-view triangulation. We use the same triangulation process as in [52], where a 3D line can be triangulated from either direct algebraic triangulation, or point/VP-guided triangulation. We refer the reader to [52] for the details of the triangulation.

We maintain a list of all the tracks that have been updated in the triangulation process and apply non-linear refinement for each updated track. In the SfM context, this refinement step shares similar functionality with the multi-view point triangulation via singular value decomposition [79].

To finish the triangulation for the new image, we merge line tracks that are connected in the matching graph. In addition, we perform a *complete* step as in [79] to collect potentially missing supports: given a line track, we consider the neighbors of the current 2D supports in the matching graph, and add them to the track if they agree with the reprojection of the 3D line. This appears to be beneficial to improve the track length under the incremental setup.

Recomputation of Endpoints. While points are compact in 3D, the spatial extent of the lines changes with their endpoints. Thus, although the infinite line may remain unchanged when the tracks are extended and merged, we need to update its endpoints to be able to correctly merge lines in the future. This is done by unprojecting endpoints onto the infinite line with Plücker coordinates [7, 52].

Retriangulation of Long 2D Lines. Long 2D lines are generally more robust than short ones since detection noise is averaged over more pixels. Thus, we force 2D lines that are longer than 100 pixels to *create* new lines through triangulation, even when they are supposed to *continue* on existing ones. This helps build more stable line tracks in the maps.

Building VP Tracks. In addition, we also maintain VP tracks to model the parallelism relations among lines. Since a VP in 3D can be mapped from a single view, its maintenance is much easier than lines. Consistency checks only involve measuring the angle of two directions. Please check Sec. A in supp. for details.

3.2 Refinement of Hybrid Structure

After each new image is registered and triangulated, our system performs local and global refinement on both the map and poses. While bundle adjustment of point features is well studied [4, 95], lines suffer comparably more from outliers and instability after triangulation. On the one hand, during refinement, these incorrect lines can potentially corrupt the entire reconstruction, especially in the early stages of SfM when only a few images are registered. As such, we need to be selective about which set of lines are added to the optimization problem. On the other hand, we do not want to prematurely filter line tracks, as many of them can become stable after more views are registered. In this section, we present several mechanisms for keeping track of the reliability of line tracks and their supports in the refinement process, without having to prematurely remove them.

Caching Inactive Supports. Due to unstable line triangulation from sparse views, a good support can easily be an outlier at an early stage due to pose perturbations in the refinement process. Deleting those supports will largely slow down the mapping process and also drop a large number of potentially good lines (see supplementary material for visualizations). Motivated by this fact, we propose to instead attach to each support an *active* label. After each refinement, we check all the supports and set their labels depending on whether it is currently an inlier (active) or an outlier (inactive). We remove the inactive ones only when a track becomes stable (has more than 10 active supports). To avoid the noisy supports to be stuck in the wrong tracks, we include each inactive support at triangulation and refinement. In this way, we keep the option for potential inlier supports to become active later in the optimization.

Propagating 3D Uncertainty from 2D Measurements. To be able to keep the noisy tracks without affecting the pose optimization, we need to correctly identify unreliable tracks. While the number of active supports makes a reasonable indicator, lines suffer more from instability due to view configurations (see Fig. 3). Therefore, we directly model the 3D line uncertainty with covariance propagation. In this paper, we assume the detection of each keypoint and line endpoint follows $\mathcal{N}(\mathbf{0}, \mathbf{1})$, while more advanced modeling [57] can be integrated easily as well.

We start by revisiting the uncertainty modeling of points, where multi-view triangulation can be formulated as a non-linear least-squares problem:

$$\boldsymbol{X}^* = \frac{1}{2} \arg\min_{\boldsymbol{X}} \sum_k \|\boldsymbol{r_k}\|^2, r_k = \Pi_k(\boldsymbol{X}) - \boldsymbol{x_k}, \tag{2}$$

where the 3D point is optimized across views w.r.t. its 2D observations $\boldsymbol{x_k}$. For such least-squares problems [3], the uncertainty can be propagated from the observation to the optimal 3D point $\boldsymbol{X}^*$ using the Jacobian J of the reprojection function. With the assumption of unit covariance, this only involves inverting the approximate Hessian $J^T J$ (refer to Sec. B.1 in supp. for details).

For lines, however, the reprojection residual (denoted as $\boldsymbol{e_k}$) is generally formulated as the endpoint-to-line distance $d_{perp}(\Pi_k(\boldsymbol{L}), \boldsymbol{l_k})$, which cannot be written in the least squares form due to the fact that the derivative of the residual over the endpoint observation depends on the optimized 3D line. This makes the 3D uncertainty intractable with the previous formulation.

We propose to tackle the problem with second-order sensitivity analysis [21], which relies on the fact that the derivatives of the non-linear objective (denoted as E) over the optimized variables is always zero at the optimal 3D line $\boldsymbol{L^*}$:

$$\frac{\partial E}{\partial \boldsymbol{L}}|_{\boldsymbol{L}=\boldsymbol{L}^*} = \sum_k \boldsymbol{e_k}^T \frac{\partial \boldsymbol{e_k}}{\partial \boldsymbol{L}}|_{\boldsymbol{L}=\boldsymbol{L}^*} = \mathbf{0} \tag{3}$$

Since the derivative of the left-hand side of Eq. (3) over the input endpoint observation $\boldsymbol{l_k}$ (*i.e.*, $\partial(\partial E/\partial \boldsymbol{L})/\partial \boldsymbol{l_k}$) always equals the zero vector at $\boldsymbol{L}^*$, we can derive a linear system that solves for the target Jacobian $\partial \boldsymbol{L^*}/\partial \boldsymbol{l_k}$. This can be used to correctly propagate the uncertainty into optimal 3D line in its Plücker form (see Sec. B in supp. material for detailed derivations). The correctness of our propagated uncertainty is supported by numerical tests with finite differences and correlation tests with map accuracy (Fig. 4).

Two-step Refinement. As previously discussed, we can combine the number of active supports and the level of uncertainty to identify reliable line tracks from unstable ones. In order to get a scale-invariant metric for the 3D uncertainty, we rescale it into pixels by multiplying it with the median value of f/d across the supporting images, where f is the focal length and d the depth of the line midpoint. Different from conventional practice in bundle adjustment, we propose to perform the refinement in two steps: an initial full hybrid bundle adjustment only including the reliable tracks, followed by a fixed-pose refinement of the unreliable tracks. After each bundle adjustment, we update the *active* label of

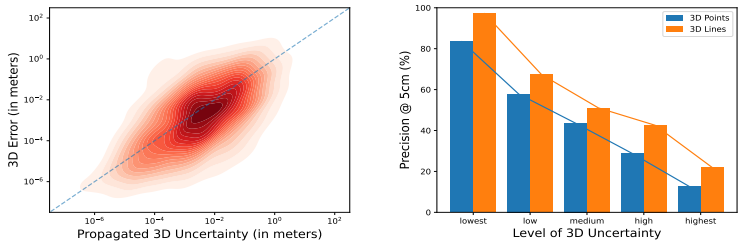


Fig. 4: Relations between the propagated uncertainty for each track and its accuracy on ETH3D [85]. Left: For each 3D feature, we plot its 3D error and uncertainty both in meters. **Right:** We report precision over each bin sorted w.r.t. to the 3D uncertainty. Points and lines with lower uncertainty tend to have higher precision.

supports and the reliability for each track. This allows us to decouple the pose optimization from unstable or incorrect line triangulations while keeping the potential for currently unreliable lines to become reliable in the future.

Integration of Structural Associations. Similarly to [52], we also incorporate structural associations between points and lines as well as lines and VPs. While this enables joint optimization with structural constraints, the point-line association residual breaks the block-wise nature of the bundle adjustment problem, leading to slower runtimes. Refer to Sec. C in the supp. mat. for a discussion.

3.3 Hybrid Registration

With the construction and maintenance of the refined hybrid maps, our system has rich information during camera registration to better pose new images with the existing 3D structure. The additional line and VP correspondences can not only help with more accurate estimation of camera poses but also enable more valid registrations on challenging images with few point correspondences.

Integration of Line and VP Correspondences. Inspired by [52], our camera registration uses a hybrid RANSAC framework [13]. Given a new image to be registered, we collect 2D-3D correspondences of points, lines, and VPs by traversing their matches with the already registered images. We employ six different minimal solvers from the combination of the hybrid correspondences, namely the conventional P3P solver [69], the hybrid point-line solvers [45, 113] (including P2P1LL, P1P2LL, P3LL), and optionally, when a VP correspondence is available, the 2-point and 1-point + 1-line solvers with one VP correspondence. The VP-based solvers are variants of the known-gravity solvers [43, 45], with the known direction tilted (check Sec. D in supp. mat for details). Following [13], the sampling probability and termination criteria for each solver depend on the corresponding inlier ratios. Both points and lines are included in the scoring and local optimization with their reprojection errors. After robust estimation, we use both the number of point and line inliers to determine whether the registration is successful. In this way, we relax the requirement of abundant inlier points with the additional line features, which are particularly common in indoor scenes.

Table 1: Structure-from-Motion results on Hypersim [74] and ETH3D [81]. We report the relative pose AUC and the percentage of valid registration within 5 cm / 5 deg after robust alignment, for both our system ("Hybrid") and COLMAP ("Point") [79] on SIFT [53] + nearest neighbor and SuperPoint [18] + SuperGlue [76].

Dataset	Point Feature	Method	AUC @ 1°/3°/5°/10° ↑				Valid Reg. ↑
Hypersim	SIFT + NN	Point	71.3	82.5	85.0	86.8	93.7%
		Hybrid	**82.1**	**86.6**	**87.6**	**88.3**	**93.9%**
	SP + SG	Point	80.1	89.5	91.6	93.2	96.7%
		Hybrid	**87.0**	**92.1**	**93.3**	**94.1**	**97.0%**
ETH3D	SIFT + NN	Point	16.2	26.7	28.1	32.1	46.4%
		Hybrid	**24.3**	**34.8**	**37.4**	**40.8**	**59.4%**
	SP + SG	Point	33.0	54.7	61.1	66.4	69.8%
		Hybrid	**37.3**	**57.9**	**63.3**	**68.8**	**75.3%**

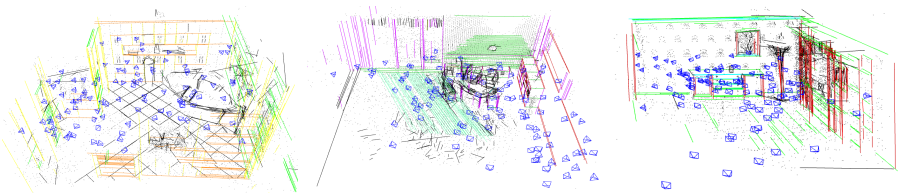


Fig. 5: Some examples of our hybrid maps on Hypersim [74]. Parallel lines from line-VP associations are colored the same.

Integration of Uncertainty Estimation. With the 3D uncertainty being propagated to each track, as described in Sec. 3.2, we can further model the reliability of different point/line correspondences from the map's perspective. It is worth noting that a 3D point/line with high 3D uncertainty can still be valuable for localization in views where its projection is stable. Thus, during registration, we model the correspondence reliability with the uncertainty of the reprojection error vector rather than the raw 3D global uncertainty. This requires an initial pose, which can be estimated by first running a few iterations of the original uncertainty-free method. After we get the reprojection uncertainty for each correspondence, we can use it as a reweighting factor in both scoring and local optimization in the robust estimation framework, which enables the stable part of the maps to contribute more to the problem. In our experiments, this uncertainty-aware mechanism achieves consistent accuracy improvement in the general localization problem on both point-alone and hybrid cases (Table 8).

4 Experiments

Implementation Details. Our system is implemented in C++ with Python bindings [35]. We use the same hyperparameters for all experiments across datasets. Parameters for points are identical to COLMAP [79] for fair comparison. For more details, refer to Sec. E in the supp. material.

Table 2: Structure-from-Motion results on PhotoTourism [86] from Image Matching Challenge 2020 [37]. We set the minimum model size to 3 following [37] for both our system ("Hybrid") and COLMAP ("Point") [79].

Point Feature	Method	AUC @ 1°/3°/5° @ N		
		$N = 5$	$N = 10$	$N = 25$
SIFT + NN	Point	**11.4** / 24.2 / 29.8	29.6 / 49.2 / 56.0	59.9 / 80.3 / 85.9
	Hybrid	11.3 / **24.6** / **30.5**	**31.3** / **51.5** / **58.1**	**63.3** / **82.4** / **87.5**
SP + SG	Point	49.0 / 75.9 / 83.8	62.2 / 84.5 / **90.3**	70.6 / 88.2 / 92.6
	Hybrid	**49.2** / **76.2** / **84.1**	**62.9** / **84.6** / 90.2	**72.3** / **89.0** / **93.1**

Table 3: Comparison between our full SfM system and the post-refinement method described in LIMAP [52].

Dataset	Method	AUC @ 1°/3°/5°/10°↑	Valid Reg. ↑
Hypersim	COLMAP [79]	71.3 / 82.5 / 85.0 / 86.8	93.7%
	COLMAP [79] → LIMAP BA [52]	78.6 / 84.2 / 86.5 / 87.3	93.8%
	Ours	**82.1** / **86.6** / **87.6** / **88.3**	**93.9%**
ETH3D	COLMAP [79]	16.2 / 26.7 / 28.1 / 32.1	46.4%
	COLMAP [79] → LIMAP BA [52]	19.2 / 28.3 / 31.1 / 33.8	47.6%
	Ours	**24.3** / **34.8** / **37.4** / **40.8**	**59.4%**

4.1 Structure-from-Motion

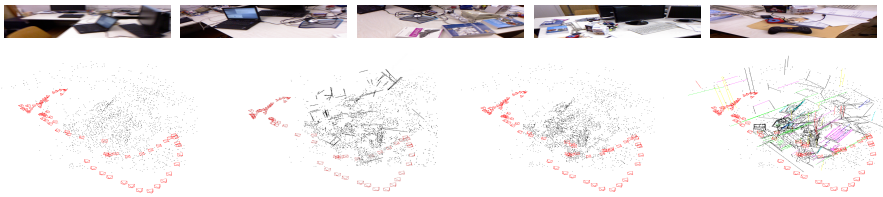
Results on Unstructured Data. We first evaluate our methods on two public datasets: Hypersim [74] and ETH3D [81]. For Hypersim, our evaluation runs on the first 8 scenes following [52]. For ETH3D, we use the training set of DSLR images (13 scenes) while resizing it to a maximum image dimension of 756. Results are shown in Table 1. Compared to the point-alone baseline COLMAP [79], our method largely improves the accuracy on highly structured indoor scenes from Hypersim, and achieves more valid registrations on ETH3D. This holds for different types of point features [18, 53, 76]. This can be attributed to the strong geometric constraints from structural features. Fig. 5 shows that our method is able to incrementally reconstruct richer maps with structural relations.

To verify that our method does not degrade the performance on scenes without abundant presence of distinctive line features, we test our SfM pipeline on the validation split (3 scenes) from the *Image Matching Benchmark 2020* [37]. Table 2 shows that despite the rich textures in the test set, our method achieves promising improvements on outdoor scenes over COLMAP [79], which remains the widely used backend in the benchmark [37]. Qualitative results in supp. further show that our method achieves reasonable reconstruction from as few as 5 images.

Comparison to Post-Refinement in [52] To better study the effectiveness of building a full SfM system, we compare our method with the hybrid post-refinement method proposed in [52] with global line triangulation. As shown in Table 3, the post-refinement method falls behind on accuracy even when most images are successfully registered, and cannot recover from the registration failure of COLMAP [79] (*e.g.* on ETH3D). Our method, on the contrary, is able to achieve more valid registrations in such challenging scenarios.

Table 4: Studies on behaviors of SLAM pipelines [62, 84] and our SfM method under different sampling rate on *fr1_desk* from [89]. While SLAM methods achieve superior results on dense frames, they suffer from low frame rates due to strong local assumptions.

Method	ATE RMSE			
	30 FPS	15 FPS	6 FPS	3 FPS
ORB-SLAM [62]	**1.27**	1.51	7.42	N/A
Structure PLP-SLAM [84]	1.64	2.36	3.88	N/A
Ours	1.39	**1.37**	**1.45**	**1.80**



ORB-SLAM [62] S. PLP-SLAM [84] COLMAP [79] Ours

Fig. 6: Map visualization with two SLAM methods [62, 84] on sequential data from [89]. Our method is able to reconstruct much richer and more complete 3D maps, especially compared to Structure PLP-SLAM [84] that also integrates line features.

Discussions on Sequential Data. We further present some studies with two popular SLAM systems: ORB-SLAM [62] and Structure PLP-SLAM [84], the latter of which also integrates line features into its pipeline. As shown in Table 4, while SLAM methods can achieve superior results on video sequences, it is constrained to sequential data and cannot deal with sparsely distributed input images. Moreover, in the qualitative results in Fig. 6, we show that our line maps are much richer and more complete compared to the top-performing SLAM counterpart [84], further highlighting the advantages of our general SfM pipeline.

4.2 Ablation Studies and More Insights

Mapping. We first study our incremental line triangulation module in terms of line reconstruction quality. Specifically, we use the proposed module to progressively triangulate 2D images with ground truth poses on Hypersim [74], which can be directly compared to the global triangulation described in [52]. Table 5 shows the results. Our method achieves comparable completeness and accuracy with the global methods. This is attributed to the carefully designed triangulation

Table 5: Quantitative results of line reconstruction on Hypersim [74]. Following the same evaluation protocols, our incremental triangulator reconstructs line maps of comparable quality with the global method in [52]. By default [52] uses top 10 matches.

Mapper	R1	R5	R10	P1	P5	P10	# lines	# supports
Global (top 10 matches) [52]	133.4	231.9	258.4	77.2	89.4	93.2	731.9	15.4 / 22.7
Global (raw matches) [52]	81.2	230.1	287.3	**71.7**	**84.7**	**90.2**	657.3	**17.9** / **22.1**
Incremental (raw matches)	**98.8**	**260.2**	**345.1**	68.9	80.5	85.3	**929.2**	14.4 / 15.6

Table 6: Ablation studies on different proposed components from mapping and refinement. Numbers are reported on Hypersim [74].

BA Ablations	AUC @ 1°/3°/5°/10° ↑
Point (COLMAP) [79]	71.3 / 82.5 / 85.0 / 86.8
+ hybrid point-line BA without caching [52]	75.9 / 83.6 / 86.0 / 87.1
+ inactive support caching	79.2 / 84.9 / 86.5 / 87.4
+ two-step refinement	80.3 / 85.6 / 86.8 / 87.8
+ retriangulation	80.8 / 85.9 / 87.0 / 87.9
+ VP associations	**81.2** / **86.1** / **87.2** / **88.0**

Table 7: Ablation studies on our registration modules in SfM. Numbers are reported on both Hypersim [74] and ETH3D [81].

Dataset	Method	AUC @ 1°/3°/5°/10°↑	Valid Reg. ↑
Hypersim	Point-based registration [25, 79]	80.8 / 85.9 / 87.0 / 87.9	93.8%
	+ hybrid registration	81.5 / 86.4 / 87.4 / 88.2	93.8%
	+ uncertainty for registration	**81.7** / **86.5** / **87.6** / **88.3**	**93.9%**
ETH3D	Point-based registration [25, 79]	19.2 / 28.0 / 30.6 / 33.0	47.6%
	+ hybrid registration	24.2 / 34.4 / 37.0 / 39.2	59.0%
	+ uncertainty for registration	**24.3** / **34.7** / **37.6** / **40.1**	**59.4%**

and maintenance strategies. In particular, thanks to the "complete" strategy inspired by [79], our incremental triangulator achieves reasonable track length while removing the need to get all the posed images beforehand.

Refinement. We further study the proposed refinement strategy together with map maintenance. We perform ablation studies on different mechanisms with the original point-alone registration from COLMAP [79]. Results in Table 6 show that each component contributes to the improvement. In particular, combining the inactive support caching and the two-step refinement method makes it possible to keep unreliable supports and tracks without corrupting the pose optimization, which speeds up the mapping process by avoiding unnecessary deletion at the early stage of track building. We include visual illustrations in Sec. F of supp.

Registration. Lastly, we study the effects of our proposed hybrid registration module with additional line features. Table 7 shows that the hybrid robust estimator consistently improves the accuracy and robustness on two different datasets, while uncertainty-aware reweighting can further improve its perfor-

Table 8: Results of our uncertainty-aware localization module on Cambridge [40] and 7Scenes [83] compared with state-of-the-art point-alone [75] and hybrid [52] methods. We report median errors (cm / deg) and recall on 3cm / 3deg and 5cm / 5deg.

Dataset	Method	Point		Point + Line	
		Med. error ↓	Recall ↑	Med. error ↓	Recall ↑
Cambridge	w/o. uncertainty	7.1 / 0.13	24.3 / 43.1	7.0 / 0.13	25.4 / 45.3
	w. uncertainty	**6.4** / **0.12**	**27.4** / **48.0**	**6.3** / **0.12**	**29.0** / **48.6**
7Scenes	w/o. uncertainty	3.1 / 1.03	51.1 / 76.0	3.1 / 1.01	52.7 / 77.7
	w. uncertainty	**2.9** / **0.95**	**55.6** / **79.0**	**2.8** / **0.95**	**56.5** / **79.5**

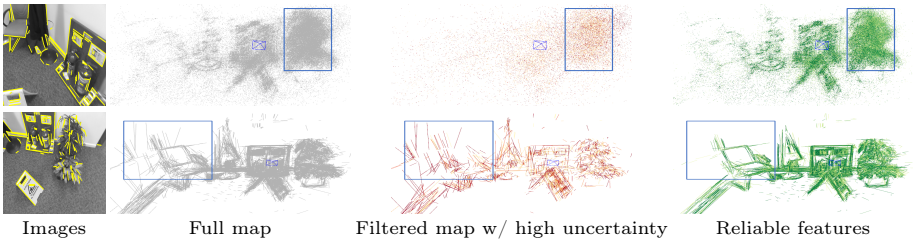


Fig. 7: Visualization of reprojected map uncertainty on *fire* from 7Scenes [83]. **First row:** 3D points. **Second row:** 3D lines. The reprojected uncertainty is a good indicator for identifying unstable reprojection of 3D maps with a given viewpoint.

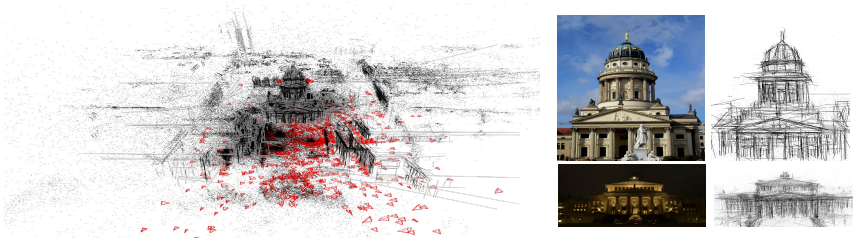


Fig. 8: Hybrid reconstruction on Gendarmenmarkt (1,463 images) from [101].

mance. Moreover, Table 8 shows results on public localization benchmarks where our uncertainty-aware localization consistently improves upon state-of-the-art practices under both point-alone [75] and hybrid [52] setup. Being able to identify the noisy features from the map (as in Fig. 7), our method increases the importance of stable 2D-3D correspondences, which can be used as a general plug-in feature in any modern localization system.

Scalability. Since our method shares a similar design as COLMAP [79], it is scalable to large-scale scenes with similar asymptotic complexity when the structural associations are disabled, while exhibiting 1-3x overhead for processing additional line features in both 2D extraction/matching and hybrid bundle adjustment. Fig. 8 shows an example of our reconstruction on 1DSfM dataset [101].

5 Conclusion

In this paper, we present a comprehensive SfM system that, in addition to points, leverages lines and their structural relations. We improve over all of the three main steps: triangulation, refinement, and registration. Experiments and ablation studies show that our method is consistently more robust and accurate compared to the widely used point-based pipeline. Additionally, our analytical uncertainty modeling benefits the localization task, as demonstrated on several public localization benchmarks. Future improvements include joint (faster) point/line detection and matching and more principled 2D uncertainty modeling. **Acknowledgements.** This work has been supported by Innosuisse funding (Grant No. 100.567 IP-ICT). V. Larsson was supported by ELLIIT and the Swedish Research Council (Grant No. 2023-05424).

References

1. Abdellali, H., Frohlich, R., Vilagos, V., Kato, Z.: L2d2: Learnable line detector and descriptor. In: International Conference on 3D Vision (3DV) (2021)
2. Agarwal, S., Furukawa, Y., Snavely, N., Simon, I., Curless, B., Seitz, S.M., Szeliski, R.: Building rome in a day. Communications of the ACM **54**(10), 105–112 (2011)
3. Agarwal, S., Mierle, K.: Ceres solver. `http://ceres-solver.org`
4. Agarwal, S., Snavely, N., Seitz, S.M., Szeliski, R.: Bundle adjustment in the large. In: ECCV (2010)
5. Åström, K., Kahl, F., Heyden, A., Berthilsson, R.: A statistical approach to structure and motion from image features. In: Advances in Pattern Recognition: Joint IAPR International Workshops SSPR'98 and SPR'98 (1998)
6. Bartoli, A., Coquerelle, M., Sturm, P.: A framework for pencil-of-points structure-from-motion. In: ECCV (2004)
7. Bartoli, A., Sturm, P.: Structure-from-motion using lines: Representation, triangulation, and bundle adjustment. Computer Vision and Image Understanding (CVIU) **100**(3), 416–441 (2005)
8. Bazin, J.C., Seo, Y., Demonceaux, C., Vasseur, P., Ikeuchi, K., Kweon, I., Pollefeys, M.: Globally optimal line clustering and vanishing point estimation in manhattan world. In: CVPR (2012)
9. Bhowmick, B., Patra, S., Chatterjee, A., Govindu, V.M., Banerjee, S.: Divide and conquer: Efficient large-scale structure from motion using graph partitioning. In: ACCV (2015)
10. Brooks, M.J., Chojnacki, W., Gawley, D., Van Den Hengel, A.: What value covariance information in estimating vision parameters? In: ICCV (2001)
11. Bui, B.T., Bui, H.H., Tran, D.T., Lee, J.H.: Representing 3d sparse map points and lines for camera relocalization. arXiv preprint arXiv:2402.18011 (2024)
12. Burnett, K., Yoon, D.J., Schoellig, A.P., Barfoot, T.D.: Radar odometry combining probabilistic estimation and unsupervised feature learning. In: Robotics: Science and Systems (RSS) (2021)
13. Camposeco, F., Cohen, A., Pollefeys, M., Sattler, T.: Hybrid camera pose estimation. In: CVPR (2018)
14. Chandraker, M., Lim, J., Kriegman, D.: Moving in stereo: Efficient structure and motion using lines. In: ICCV (2009)
15. Chum, O., Matas, J., Kittler, J.: Locally optimized ransac. In: Joint Pattern Recognition Symposium (2003)
16. Crandall, D., Owens, A., Snavely, N., Huttenlocher, D.: Discrete-continuous optimization for large-scale structure from motion. In: CVPR (2011)
17. Dellaert, F., Seitz, S.M., Thorpe, C.E., Thrun, S.: Structure from motion without correspondence. In: CVPR (2000)
18. DeTone, D., Malisiewicz, T., Rabinovich, A.: Superpoint: Self-supervised interest point detection and description. In: Computer Vision and Pattern Recognition Workshops (CVPRW) (2018)
19. Dusmanu, M., Miksik, O., Schönberger, J.L., Pollefeys, M.: Cross-Descriptor Visual Localization and Mapping. In: ICCV (2021)
20. Dusmanu, M., Schönberger, J.L., Pollefeys, M.: Multi-View Optimization of Local Feature Geometry. In: ECCV (2020)
21. Fiacco, A.V., Ishizuka, Y.: Sensitivity and stability analysis for nonlinear programming. Annals of Operations Research **27**(1), 215–235 (1990)

22. Förstner, W., Gülch, E.: A fast operator for detection and precise location of distinct points, corners and centres of circular features. In: Proc. ISPRS intercommission conference on fast processing of photogrammetric data (1987)
23. Förstner, W., Wrobel, B.P.: Photogrammetric computer vision (2016)
24. Frahm, J.M., Fite-Georgel, P., Gallup, D., Johnson, T., Raguram, R., Wu, C., Jen, Y.H., Dunn, E., Clipp, B., Lazebnik, S., et al.: Building rome on a cloudless day. In: ECCV (2010)
25. Gao, X.S., Hou, X.R., Tang, J., Cheng, H.F.: Complete solution classification for the perspective-three-point problem. IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI) **25**(8), 930–943 (2003)
26. Germain, H., Bourmaud, G., Lepetit, V.: S2dnet: Learning accurate correspondences for sparse-to-dense feature matching. In: ECCV (2020)
27. Gomez-Ojeda, R., Gonzalez-Jimenez, J.: Robust stereo visual odometry through a probabilistic combination of points and line segments. In: ICRA (2016)
28. Gomez-Ojeda, R., Moreno, F.A., Zuniga-Noël, D., Scaramuzza, D., Gonzalez-Jimenez, J.: Pl-slam: A stereo slam system through the combination of points and line segments. IEEE Transactions on Robotics **35**(3), 734–746 (2019)
29. Hartley, R.I., Sturm, P.: Triangulation. Computer Vision and Image Understanding (CVIU) **68**(2), 146–157 (1997)
30. He, X., Sun, J., Wang, Y., Peng, S., Huang, Q., Bao, H., Zhou, X.: Detector-free structure from motion. arXiv preprint arXiv:2306.15669 (2023)
31. He, Y., Zhao, J., Guo, Y., He, W., Yuan, K.: Pl-vio: Tightly-coupled monocular visual–inertial odometry using point and line features. Sensors **18**(4), 1159 (2018)
32. Hofer, M., Maurer, M., Bischof, H.: Line3d: Efficient 3d scene abstraction for the built environment. In: German Conference on Pattern Recognition (2015)
33. Holynski, A., Geraghty, D., Frahm, J.M., Sweeney, C., Szeliski, R.: Reducing drift in structure from motion using extended features. In: International Conference on 3D Vision (3DV) (2020)
34. Huang, S., Qin, F., Xiong, P., Ding, N., He, Y., Liu, X.: Tp-lsd: Tri-points based line segment detector. In: ECCV (2020)
35. Jakob, W., Rhinelander, J., Moldovan, D.: pybind11 – seamless operability between c++11 and python. `https://github.com/pybind/pybind11`
36. Jiang, N., Cui, Z., Tan, P.: A global linear method for camera pose registration. In: ICCV (2013)
37. Jin, Y., Mishkin, D., Mishchuk, A., Matas, J., Fua, P., Yi, K.M., Trulls, E.: Image matching across wide baselines: From paper to practice. IJCV **129**(2), 517–547 (2021)
38. Kanatani, K.: For geometric inference from images, what kind of statistical model is necessary? Systems and Computers in Japan **35**(6), 1–9 (2004)
39. Kanazawa, Y., Kanatani, K.: Do we really have to consider covariance matrices for image feature points? Electronics and communications in Japan (part III: Fundamental electronic science) **86**(1), 1–10 (2003)
40. Kendall, A., Grimes, M., Cipolla, R.: PoseNet: A convolutional network for real-time 6-DoF camera relocalization. In: ICCV (2015)
41. Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 3d gaussian splatting for real-time radiance field rendering. ACM Transactions on Graphics (ToG) **42**(4), 1–14 (2023)
42. Kuhn, A., Sormann, C., Rossi, M., Erdler, O., Fraundorfer, F.: Deepc-mvs: Deep confidence prediction for multi-view stereo reconstruction. In: International Conference on 3D Vision (3DV) (2020)

43. Kukelova, Z., Bujnak, M., Pajdla, T.: Closed-form solutions to minimal absolute pose problems with known vertical direction. In: ACCV (2010)
44. Kushal, A., Agarwal, S.: Visibility based preconditioning for bundle adjustment. In: CVPR (2012)
45. Larsson, V.: PoseLib - Minimal Solvers for Camera Pose Estimation. `https://github.com/vlarsson/PoseLib`, `https://github.com/vlarsson/PoseLib`
46. Lebeda, K., Matas, J., Chum, O.: Fixing the locally optimized ransac–full experimental evaluation. In: BMVC (2012)
47. Li, H., Zhao, J., Bazin, J.C., Chen, W., Liu, Z., Liu, Y.H.: Quasi-globally optimal and efficient vanishing point estimation in manhattan world. In: ICCV (2019)
48. Li, Z., Müller, T., Evans, A., Taylor, R.H., Unberath, M., Liu, M.Y., Lin, C.H.: Neuralangelo: High-fidelity neural surface reconstruction. In: CVPR (2023)
49. Lim, H., Jeon, J., Myung, H.: Uv-slam: Unconstrained line-based slam using vanishing points for structural mapping. IEEE Robotics and Automation Letters (RA-L) **7**(2), 1518–1525 (2022)
50. Lim, H., Kim, Y., Jung, K., Hu, S., Myung, H.: Avoiding degeneracy for monocular visual slam with point and line features. In: ICRA (2021)
51. Lindenberger, P., Sarlin, P.E., Larsson, V., Pollefeys, M.: Pixel-perfect structure-from-motion with featuremetric refinement. In: ICCV (2021)
52. Liu, S., Yu, Y., Pautrat, R., Pollefeys, M., Larsson, V.: 3d line mapping revisited. In: CVPR (2023)
53. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. IJCV **60**(2), 91–110 (2004)
54. Lu, F., Hartley, R.: A fast optimal algorithm for l 2 triangulation. In: ACCV (2007)
55. Marzorati, D., Matteucci, M., Migliore, D., Sorrenti, D.G.: Integration of 3d lines and points in 6dof visual slam by uncertain projective geometry. In: EMCR (2007)
56. Mateus, A., Tahri, O., Aguiar, A.P., Lima, P.U., Miraldo, P.: On incremental structure from motion using lines. IEEE Transactions on Robotics **38**(1), 391–406 (2021)
57. Meidow, J., Beder, C., Förstner, W.: Reasoning with uncertain points, straight lines, and straight line segments in 2d. ISPRS Journal of Photogrammetry and Remote Sensing **64**(2), 125–139 (2009)
58. Micusik, B., Wildenauer, H.: Structure from motion with line segments under relaxed endpoint constraints. IJCV **124**(1), 65–79 (2017)
59. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. In: ECCV (2020)
60. Moulon, P., Monasse, P., Perrot, R., Marlet, R.: OpenMVG: Open multiple view geometry. In: International Workshop on Reproducible Research in Pattern Recognition (2016)
61. Muhle, D., Koestler, L., Jatavallabhula, K.M., Cremers, D.: Learning correspondence uncertainty via differentiable nonlinear least squares. In: CVPR (2023)
62. Mur-Artal, R., Montiel, J.M.M., Tardos, J.D.: Orb-slam: a versatile and accurate monocular slam system. IEEE Transactions on Robotics **31**(5), 1147–1163 (2015)
63. Nistér, D.: Preemptive ransac for live structure and motion estimation. Machine Vision and Applications **16**(5), 321–329 (2005)
64. Nurutdinova, I., Fitzgibbon, A.: Towards pointless structure from motion: 3d reconstruction and camera parameters from general 3d curves. In: ICCV (2015)
65. Pautrat, R., Lin, J.T., Larsson, V., Oswald, M.R., Pollefeys, M.: Sold2: Self-supervised occlusion-aware line description and detection. In: CVPR (2021)

66. Pautrat, R., Liu, S., Hruby, P., Pollefeys, M., Barath, D.: Vanishing point estimation in uncalibrated images with prior gravity direction. In: ICCV (2023)
67. Pautrat, R., Suárez, I., Yu, Y., Pollefeys, M., Larsson, V.: Gluestick: Robust image matching by sticking points and lines together. In: ICCV (2023)
68. Pautrat, R., Barath, D., Larsson, V., Oswald, M.R., Pollefeys, M.: Deeplsd: Line segment detection and refinement with deep image gradients. In: CVPR (2023)
69. Persson, M., Nordberg, K.: Lambda twist: An accurate fast robust perspective three point (p3p) solver. In: ECCV (2018)
70. Poggi, M., Mattoccia, S.: Learning from scratch a confidence measure. In: BMVC (2016)
71. Pumarola, A., Vakhitov, A., Agudo, A., Sanfeliu, A., Moreno-Noguer, F.: Pl-slam: Real-time monocular visual slam with points and lines. In: ICRA (2017)
72. Qian, G., Chellappa, R.: Structure from motion using sequential monte carlo methods. IJCV **59**, 5–31 (2004)
73. Qian, Y., Elder, J.H.: A reliable online method for joint estimation of focal length and camera rotation. In: ECCV (2022)
74. Roberts, M., Ramapuram, J., Ranjan, A., Kumar, A., Bautista, M.A., Paczan, N., Webb, R., Susskind, J.M.: Hypersim: A photorealistic synthetic dataset for holistic indoor scene understanding. In: ICCV (2021)
75. Sarlin, P.E.: Visual localization made easy with hloc. `https://github.com/cvg/Hierarchical-Localization/`
76. Sarlin, P.E., DeTone, D., Malisiewicz, T., Rabinovich, A.: Superglue: Learning feature matching with graph neural networks. In: CVPR (2020)
77. Sarlin, P.E., Dusmanu, M., Schönberger, J.L., Speciale, P., Gruber, L., Larsson, V., Miksik, O., Pollefeys, M.: LaMAR: Benchmarking Localization and Mapping for Augmented Reality. In: ECCV (2022)
78. Schindler, G., Krishnamurthy, P., Dellaert, F.: Line-based structure from motion for urban environments. In: International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT) (2006)
79. Schonberger, J.L., Frahm, J.M.: Structure-from-motion revisited. In: CVPR (2016)
80. Schöps, T., Engel, J., Cremers, D.: Semi-dense visual odometry for ar on a smartphone. In: International Symposium on Mixed and Augmented Reality (ISMAR) (2014)
81. Schops, T., Schonberger, J.L., Galliani, S., Sattler, T., Schindler, K., Pollefeys, M., Geiger, A.: A multi-view stereo benchmark with high-resolution images and multi-camera videos. In: CVPR (2017)
82. Seki, A., Pollefeys, M.: Patch based confidence prediction for dense disparity map. In: BMVC (2016)
83. Shotton, J., Glocker, B., Zach, C., Izadi, S., Criminisi, A., Fitzgibbon, A.: Scene coordinate regression forests for camera relocalization in RGB-D images. In: CVPR (2013)
84. Shu, F., Wang, J., Pagani, A., Stricker, D.: Structure plp-slam: Efficient sparse mapping and localization using point, line and plane for monocular, rgb-d and stereo cameras. In: ICRA (2023)
85. Sinha, S.N., Steedly, D., Szeliski, R.: A multi-stage linear approach to structure from motion. In: Trends and Topics in Computer Vision: ECCV 2010 Workshops, Heraklion, Crete, Greece, September 10-11, 2010, Revised Selected Papers, Part II 11 (2012)
86. Snavely, N., Seitz, S.M., Szeliski, R.: Photo tourism: exploring photo collections in 3d. In: ACM SIGGRAPH (2006)

87. Steedly, D., Essa, I.A., Dellaert, F.: Spectral partitioning for structure from motion. In: ICCV (2003)
88. Steele, R.M., Jaynes, C.: Feature uncertainty arising from covariant image noise. In: CVPR (2005)
89. Sturm, J., Engelhard, N., Endres, F., Burgard, W., Cremers, D.: A benchmark for the evaluation of rgb-d slam systems. In: IROS (2012)
90. Sweeney, C.: Theia multiview geometry library: Tutorial & reference. `http://theia-sfm.org`
91. Sweeney, C., Sattler, T., Hollerer, T., Turk, M., Pollefeys, M.: Optimizing the viewing graph for structure-from-motion. In: ICCV (2015)
92. Tang, C., Tan, P.: Ba-net: Dense bundle adjustment network. In: International Conference on Learning Representations (ICLR) (2019)
93. Taylor, C.J., Kriegman, D.J.: Structure and motion from line segments in multiple images. IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI) **17**(11), 1021–1032 (1995)
94. Toldo, R., Fusiello, A.: Robust multiple structures estimation with j-linkage. In: ECCV (2008)
95. Triggs, B., McLauchlan, P.F., Hartley, R.I., Fitzgibbon, A.W.: Bundle adjustment—a modern synthesis. In: Vision Algorithms: Theory and Practice: International Workshop on Vision Algorithms Corfu, Greece, September 21–22, 1999 Proceedings (2000)
96. Von Gioi, R.G., Jakubowicz, J., Morel, J.M., Randall, G.: Lsd: A fast line segment detector with a false detection control. IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI) **32**(4), 722–732 (2008)
97. Wang, J., Karaev, N., Rupprecht, C., Novotny, D.: Visual geometry grounded deep structure from motion. In: CVPR (2024)
98. Wang, J., Rupprecht, C., Novotny, D.: Posediffusion: Solving pose estimation via diffusion-aided bundle adjustment. In: ICCV (2023)
99. Wang, S., Leroy, V., Cabon, Y., Chidlovskii, B., Revaud, J.: Dust3r: Geometric 3d vision made easy. In: CVPR (2024)
100. Wei, X., Huang, J., Ma, X.: Real-time monocular visual slam by combining points and lines. In: IEEE International Conference on Multimedia and Expo (ICME) (2019)
101. Wilson, K., Snavely, N.: Robust global translations with 1dsfm. In: ECCV (2014)
102. Wu, C.: Visualsfm: A visual structure from motion system. http://www. cs. washington. edu/homes/ccwu/vsfm (2011)
103. Wu, C.: Towards linear-time incremental structure from motion. In: International Conference on 3D Vision (3DV) (2013)
104. Xiao, Y., Xue, N., Wu, T., Xia, G.S.: Level-s2 fm: Structure from motion on neural level set of implicit surfaces. In: CVPR (2023)
105. Xue, N., Wu, T., Bai, S., Wang, F., Xia, G.S., Zhang, L., Torr, P.H.: Holistically-attracted wireframe parsing. In: CVPR (2020)
106. Yan, J., Zheng, Y., Yang, J., Mihaylova, L., Yuan, W., Gu, F.: Plpf-vslam: An indoor visual slam with adaptive fusion of point-line-plane features. Journal of Field Robotics (2023)
107. Yu, Z., Peng, S., Niemeyer, M., Sattler, T., Geiger, A.: Monosdf: Exploring monocular geometric cues for neural implicit surface reconstruction. In: NeurIPS (2022)
108. Zeisl, B., Georgel, P.F., Schweiger, F., Steinbach, E.G., Navab, N., Munich, G.: Estimation of location uncertainty for scale invariant features points. In: BMVC (2009)

109. Zhang, H., Grießbach, D., Wohlfeil, J., Börner, A.: Uncertainty model for template feature matching. In: Image and Video Technology: 8th Pacific-Rim Symposium, PSIVT 2017, Wuhan, China, November 20-24, 2017, Revised Selected Papers 8 (2018)
110. Zhang, J.Y., Lin, A., Kumar, M., Yang, T.H., Ramanan, D., Tulsiani, S.: Cameras as rays: Pose estimation via ray diffusion. In: International Conference on Learning Representations (ICLR) (2024)
111. Zhang, L., Lu, H., Hu, X., Koch, R.: Vanishing point estimation and line classification in a manhattan world with a unifying camera model. IJCV **117** (2015)
112. Zhao, W., Liu, S., Wei, Y., Guo, H., Liu, Y.J.: A confidence-based iterative solver of depths and surface normals for deep multi-view stereo. In: ICCV (2021)
113. Zhou, L., Ye, J., Kaess, M.: A stable algebraic camera pose estimation for minimal configurations of 2d/3d point and line correspondences. In: ACCV (2018)
114. Zuo, X., Xie, X., Liu, Y., Huang, G.: Robust visual slam with point and line features. In: IROS (2017)