Gaussian Uncertainty-Driven Multi-Model Fitting with Graph Neural Network

Ligang Zhang[1,2,3,4,*], **Jun Li**[1,2,3,4,*], **Qiming Li**[1,2,3,4,†]

[1]Quanzhou Institute of Equipment Manufacturing, Haixi Institute, Chinese Academy of Sciences, Quanzhou, 362216, China
[2]Fujian College, University of Chinese Academy of Sciences, Fuzhou, 350000, China
[3]Fujian Institute of Research on the Structure of Matter, Chinese Academy of Sciences, Fuzhou, 350000, China
[4]School of Advanced Manufacturing, Fuzhou University, Quanzhou, 362200, China
{qimingli, junli}@fjirsm.ac.cn, 238527306@fzu.edu.cn

Abstract

Multi-model fitting is fundamental for robust geometric estimation in computer vision. However, recent deep learning methods enable parallel model detection but rely on simple architectures that inadequately model spatial relationships. Moreover, current methods typically generate hypotheses only through minimal solvers on randomly sampled points, thus failing to explore the full diversity of the solution space. To address these limitations, we propose a novel Jacobian-based Gaussian uncertainty modeling framework, which analytically propagates covariance through geometric transformations and enables efficient expansion of the hypothesis space with strong theoretical guarantees. We further introduce a Gaussian Hypothesis Generation Network (GHG-Net) to learn global parameter distributions, enabling the generation of diverse and geometrically valid hypotheses. Additionally, our network captures spatial relationships among observations by employing a dynamic graph neural network with a multi-head attention mechanism. This yields more accurate sample and inlier weights, significantly improving the quality of hypothesis generation. Extensive experiments on three representative geometric estimation tasks (i.e. vanishing point detection, fundamental matrix estimation, and homography estimation) demonstrate that our method achieves new state-of-the-art accuracy and stability, while maintaining high computational efficiency.

Introduction

Geometric model fitting constitutes a cornerstone of computer vision, enabling the interpretation of complex visual data through parametric representations that capture underlying scene structure. This process transforms high-dimensional noisy observations into meaningful geometric primitives—vanishing points for architectural understanding, homography matrices for planar surface analysis, and fundamental matrices for epipolar geometry estimation. These geometric abstractions form the computational foundation for diverse applications including large-scale 3D reconstruction (Heinly et al. 2015; Schonberger and Frahm 2016; Wandt and Rosenhahn 2019), autonomous navigation (Kluger et al. 2020a; Mur-Artal and Tardós 2017; Sattler, Leibe, and Kobbelt 2016), and immersive augmented reality systems (Abu Alhaija et al. 2018; Brachmann and Rother 2018). Modern geometric fitting pipelines follow a two-stage paradigm: feature extraction algorithms (including SIFT (Lowe 2004), LSD (Von Gioi et al. 2008), and DeepLSD (Pautrat et al. 2023)) generate candidate observations that inevitably contain noise and false detections; robust parameter estimation that must discriminate between valid measurements (inliers) and erroneous data (outliers). While single-model estimation is well-developed, multi-model scenarios remain fundamentally challenging.

Multi-model fitting has evolved through distinct methodological paradigms over past few decades. Early solutions employ sequential strategies, iteratively applying robust estimators such as RANSAC (Fischler and Bolles 1981) to identify individual models while progressively removing associated observations. Despite their conceptual simplicity, sequential approaches suffer from error propagation and model interference issues. Later methods adopt more sophisticated strategies that leverage clustering and optimization methods to simultaneously assign observations to multiple model hypotheses or outlier categories. This paradigm shift has led to many influential algorithms (Magri, Fusiello et al. 2015; Barath et al. 2023), each contributing novel mechanisms for hypothesis management and inlier-outlier discrimination across complex geometric configurations.

Recent advances have introduced deep learning methods (Brachmann and Rother 2019; Cavalli, Pollefeys, and Barath 2022; Cavalli et al. 2023; Wei et al. 2023; Wei, Matas, and Barath 2023; Kluger et al. 2020b; Kluger and Rosenhahn 2024) into robust geometric model fitting methods. Despite improved accuracy, these methods face three major limitations. First, they lack representational deficiency, as they process observations independently without capturing essential spatial relationships and geometric dependencies. Second, their reliance on deterministic sampling limits hypothesis diversity and exploration of the solution space. Third, they lack uncertainty quantification, generating point estimates without characterizing parameter distributions or propagating uncertainty from observation to parameter space. These limitations significantly impair performance in complex scenarios with multiple models, high noise levels, or ambiguous geometric configurations. Therefore, they call for a unified framework that enhances both

*These authors contributed equally.
†Corresponding author.

representational capacity and principled hypothesis generation through structured geometric reasoning.

In this paper, we propose a robust multi-model fitting method that addresses the above mentioned limitations in hypothesis generation and evaluation through principled Gaussian uncertainty modeling. We develop a comprehensive uncertainty quantification framework that propagates information from the observation space to the parameter space through analytical Jacobian analysis. This methodology characterizes model hypothesis uncertainty as Gaussian distributions in the parameter space, enabling diverse hypothesis generation via mathematically grounded sampling strategies. To complement this local exploration, we further introduce the Gaussian Hypothesis Generation Network (GHG-Net), which learns global parameter distributions from training data to ensure comprehensive solution space coverage. To further enhance robustness, we incorporate dynamic graph neural networks to capture complex spatial relationships among observations. Unlike prior methods that treat observations independently, our method constructs adaptive graph structures based on feature similarity and spatial proximity, where nodes represent observations and edges encode geometric relationships. Through multi-head attention within our graph architecture, our method captures both local dependencies and global contextual information, enabling enhanced understanding of underlying geometric structures. This graph-based representation significantly improves inlier-outlier discrimination across multiple model instances, especially under severe noise. By combining principled uncertainty modeling with graph-based learning, our method robustly handles complex multi-model scenarios of varying geometric complexity, addressing both hypothesis diversity and contextual understanding.

Our method preserves the computational advantages of parallel processing while substantially enhancing representational capacity and hypothesis generation capabilities. As shown in Fig. 1 (b) and Fig. 1 (c), benefiting from the effective modeling of spatial relationships among observations through graph structures, we obtain more accurate sample and inlier weights, thereby generating more precise initial hypotheses. Subsequently, through Gaussian uncertainty modeling based on Jacobian matrix analysis and GHG-Net that learns global parameter distributions, we construct a richer and more diverse hypothesis pool, significantly increasing the opportunity to find optimal geometric models. As shown in Fig. 1 (d), compared to the current state-of-the-art PARSAC method, our method can get more accurate geometric model in complex scenes while maintaining the computational efficiency of parallel processing.

In summary, our main contributions are as follows:

- We replace independent point-wise processing with dynamic graph construction and multi-head attention mechanisms, enabling explicit modeling of spatial relationships and geometric dependencies among observations.
- We apply analytical Jacobian matrices to three geometric tasks (vanishing point, homography, and fundamental matrix estimation), enabling uncertainty propagation via the Fisher Information Matrix for principled Gaussian hypothesis generation under geometric constraints.
- We propose the GHG-Net that learns global parameter distributions, complemented by local Jacobian-based uncertainty sampling, forming a dual-level framework that systematically balances global exploration with local refinement.

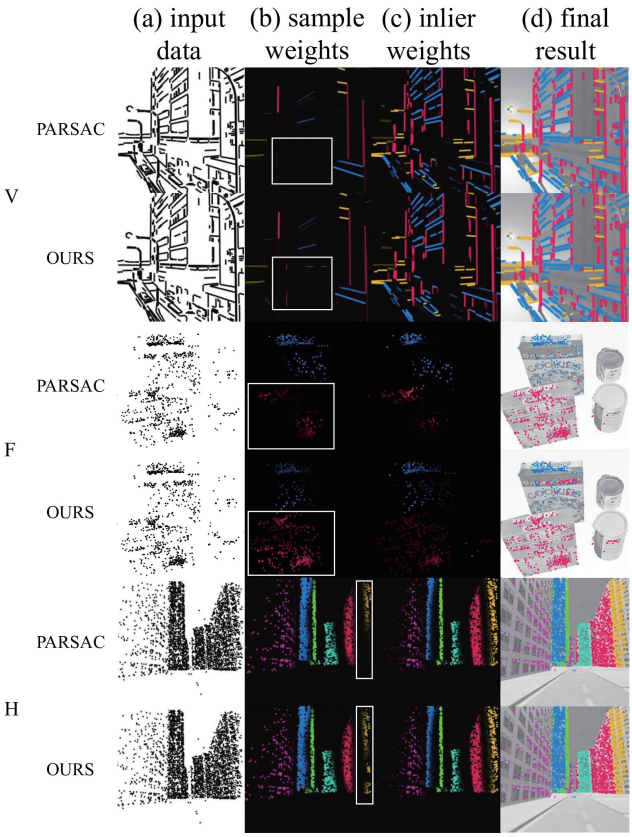

Figure 1: We apply our method to vanishing point estimation (V, top), fundamental matrix estimation (F, middle), and homography matrix estimation (H, bottom), and we also have current state-of-the-art results in deep learning PARSAC for comparison (differences are highlighted in white boxes).

Related Work

Multi-Model Fitting

Robust model fitting aims to estimate geometric models from outlier-contaminated observations. The foundational RANSAC (Fischler and Bolles 1981) algorithm generates hypotheses through minimal set sampling and selects models with maximum consensus. Although effective for single-model scenarios, RANSAC struggles when multiple model instances coexist in the data. Early multi-model fitting methods employ sequential strategies. Sequential RANSAC (Vincent and Laganiere 2001) iteratively applies RANSAC while removing identified inliers, but suffers from error propagation as subsequent models are influenced by earlier decisions. PEARL (Isack and Boykov 2012) introduces energy-based optimization using global mixed-integer cost func-

tions, while Multi-X (Barath and Matas 2018) extends this framework to multi-class scenarios. Progressive-X (Barath and Matas 2019) improves efficiency by interleaving sampling and optimization steps, with its successor Progressive-X+ (Barath et al. 2023) incorporating preference analysis and multi-model assignment capabilities. To improve efficiency, several strategies have been developed. TRESAC (Guo et al. 2022) utilizes triplet relationships to enforce spatial consistency, while SPGSC (Guo et al. 2024) guides sampling through second-order proximity measures. LSC (Xiao et al. 2024) maintains latent semantic consistency between data points and hypotheses to improve the quality of generated hypotheses. Clustering-based methods have also gained prominence. J-Linkage (Toldo and Fusiello 2008) and T-Linkage (Magri and Fusiello 2014) employ agglomerative clustering based on preference sets, while RPA (Magri, Fusiello et al. 2015) utilizes spectral clustering techniques. Graph-based methods are widely used to model complex geometric relationships. Methods like GMSF (Xiao et al. 2016) leverage both local and global information through graph structures, while CBG (Lin et al. 2022a) utilizes bipartite graphs to analyze correlations between data points and model hypotheses. Hypergraph representations in MSHF (Wang et al. 2018), SWS+NCut (Purkait et al. 2016), and HOMF (Lin et al. 2019) characterize higher-order geometric dependencies. Recent developments include specialized methods like Fast-CP (Ozbay, Camps, and Sznaier 2022) for fundamental matrix estimation. Matteo Farina et al. (Farina et al. 2023) pioneer the application of quantum computing for multi-model fitting, but their work is conducted under restrictive underlying assumptions.

The integration of deep learning introduces new paradigms in robust fitting. Neural-guided methods (Brachmann and Rother 2019; Cavalli, Pollefeys, and Barath 2022; Cavalli et al. 2023; Wei et al. 2023; Wei, Matas, and Barath 2023; Kluger et al. 2020b; Kluger and Rosenhahn 2024) leverage learned weights to suppress outlier influence. For multi-model fitting, CONSAC (Kluger et al. 2020b) pioneers neural-guided sequential discovery, while PARSAC (Kluger and Rosenhahn 2024) achieves parallel processing through decoupled model instance discovery, significantly reducing computational overhead. Despite recent progress, current learning-based multi-model fitting methods rely on conventional convolutional architectures, limiting their ability to model spatial relationships essential for geometric understanding. Building on current methods, we propose a novel method that bridges this gap by integrating dynamic graph construction and multi-head attention, along with analytical Jacobian-based uncertainty propagation to improve hypothesis generation.

Vanishing Point Estimation

Vanishing point is a classic example of multi-model fitting in geometric computer vision. Specialized methods (Tardif 2009; Barinova et al. 2010; Wildenauer and Hanbury 2012; Antunes and Barreto 2013; Lezama et al. 2014; Zhai, Workman, and Jacobs 2016; Zhou et al. 2019; Simon, Fond, and Berger 2018; Wu et al. 2021; Liu, Zhou, and Zhao 2021; Lin et al. 2022b) have leveraged domain-specific geometric constraints and tailored network architectures to achieve superior performance within constrained scenarios. These domain-specific methods achieve high accuracy by incorporating geometric priors and architectural constraints tailored to vanishing point characteristics. However, their specialized nature limits transferability to other geometric fitting problems. Conversely, general-purpose robust estimation frameworks such as CONSAC and PARSAC offer algorithmic flexibility across diverse geometric problems through universal sampling strategies and consensus mechanisms. Although broadly applicable, these methods rely on conventional convolutional architectures that treat observations independently, limiting their ability to capture the spatial dependencies inherent in geometric data. Our method addresses this limitation while maintaining the general-purpose characteristics of multi-model fitting algorithms.

Method

To explicitly capture complex spatial relationships among observations, we introduce a graph neural network, which constructs dynamic graphs, where observations serve as nodes and edges encode spatial proximity and geometric compatibility. This graph representation enables message passing between spatially related observations, allowing the network to leverage geometric context essential for robust multi-model fitting. The overall architecture is illustrated in Fig. 2. We stack observations $\mathrm{x} \in \mathcal{X}$ into tensors of size $N \times 1 \times D$, where D is fixed at 4. x is a line segment parametrised via centroid (x, y), length l, and angle α for vanishing points. x is a point correspondence pair (x_1, y_1, x_2, y_2) for fundamental matrices and homographies. To extract meaningful features from observations, we first apply an initial feature embedding layer that projects the raw observations into a higher-dimensional feature space. Given these embedded features X, we construct dynamic graphs by computing pairwise similarities S_{ij} and selecting k-nearest neighbors to form symmetric adjacency matrices. Multi-head graph attention captures contextual relationships, the local pathway uses graph convolution while the global pathway employs attention mechanisms. Cross-attention fusion combines both pathways through bidirectional attention and adaptive gating. The fused features then undergo enhanced spatial feature aggregation with multi-scale pooling, channel attention, and spatial attention mechanisms, before generating dual outputs for inlier weights and sample weights to enable robust hypothesis generation. Please refer to the extended version for implementation details.

Loss Function. We adopt the optimization framework as PARSAC, which formulates the objective as minimizing the expected task loss:

$$L(w) = E_{M \sim p(M|\mathcal{X};w)}[\ell(M)], \tag{1}$$

where M represents the estimated model instances, and $\ell(M)$ measures model quality through task-specific metrics.

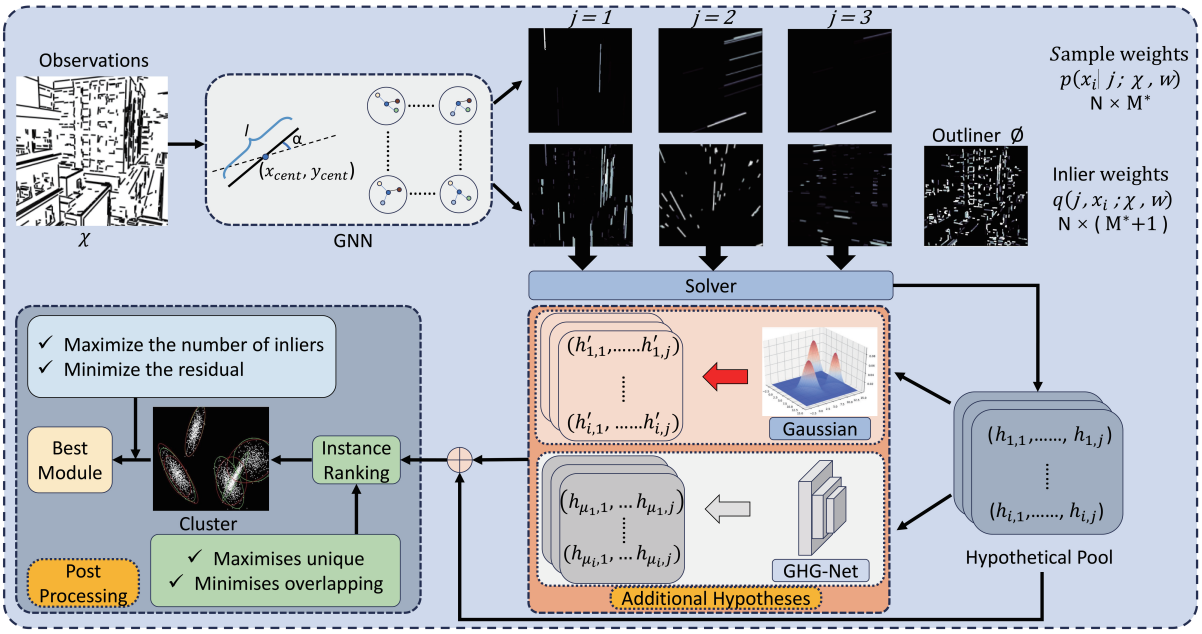

Figure 2: Overview architecture: The system processes input observations through a GNN-based network to generate sample and inlier weights. The hypothesis pool is enriched through two complementary pathways: Gaussian uncertainty modeling that analytically expands hypotheses around current estimates, and GHG-Net that learns global parameter distributions from training data. This dual-pathway design balances local precision with global exploration for comprehensive parameter space coverage.

Gaussian Uncertainty Modeling

Geometric parameter estimation often exhibits complex correlations and anisotropic uncertainty. Traditional deterministic approaches struggle with such ill-posed problems where minor data perturbations cause substantial parameter variations. We propose analytical Gaussian uncertainty modeling that captures parameter dependencies through closed-form Jacobian matrices.

Jacobian-based Uncertainty Quantification. Our key insight is that different hypotheses exhibit varying degrees of reliability and uncertainty. By quantifying this uncertainty, we can more effectively guide the hypothesis generation process and improve model selection. For each geometric problem, we derive analytical Jacobian matrices J_θ by computing derivatives of specific residual functions with respect to model parameters. For vanishing point estimation, we use the angular residual:

$$r_V = 1 - \cos\alpha. \tag{2}$$

For fundamental matrix estimation, we employ the Sampson error:

$$r_F = \frac{(p_b^T F p_a)^2}{(Fp_a)_1^2 + (Fp_a)_2^2 + (F^T p_b)_1^2 + (F^T p_b)_2^2}, \tag{3}$$

where F is the fundamental matrix, and (p_a, p_b) is point correspondence pair.
For homography estimation, we utilize the symmetric transfer error:

$$r_H = d(p_a, H^{-1}p_b)^2 + d(p_b, Hp_a)^2, \tag{4}$$

where H is the homography matrix. These Jacobians establish the critical relationship between parameter variations and corresponding changes in residual values. The Fisher Information Matrix provides uncertainty bounds:

$$I(\theta) = \sum_{i=1}^{N} \frac{1}{\sigma_i^2} J_{\theta,i}^T J_{\theta,i}. \tag{5}$$

Parameter covariance is estimated as $\Sigma_\theta = I(\theta)^{-1}$, capturing both uncertainty magnitude and correlations. It enables adaptive hypothesis expansion in regions of high uncertainty, concentrating computational resources where parameter estimation is most challenging. Unlike traditional methods that treat all hypotheses equally, our Jacobian-based uncertainty quantification creates a structured hypothesis space that reflects the underlying geometry of the problem, significantly improving both efficiency and accuracy in multi-model fitting. Please refer to the extended version for implementation details.

Gaussian Hypothesis Generation. We construct multivariate Gaussian distributions $P(\theta) \sim \mathcal{N}(\hat{\theta}, \Sigma_\theta)$ and generate samples via Cholesky decomposition:

$$\theta_{gen} = \hat{\theta} + L_{math} z, \tag{6}$$

where L_{math} is the Cholesky factor of $\Sigma_\theta = L_{math} L_{math}^T$, z follows a standard multivariate normal distribution, denoted as $z \sim \mathcal{N}(0, I)$. This uncertainty-aware sampling adaptively explores the parameter space according to estimated uncertainty, preserves correlations, and reduces invalid hypotheses compared to uniform sampling.

Gaussian Hypothesis Generation Network

While analytical Gaussian modeling provides locally precise hypotheses around current parameter estimates through Fisher information analysis, it inherently suffers from limited exploration capability—the generated hypotheses cluster around existing solutions and may miss distant but valid parameter regions. This limitation becomes particularly problematic in multi-model scenarios where multiple distinct solutions coexist in the parameter space. To address this exploration-exploitation trade-off, we propose a GHG-Net to learn global parameter distributions from training data across diverse geometric configurations. By capturing broad statistical patterns of valid geometric parameters, GHG-Net generates hypotheses that explore previously unvisited regions of the parameter space, creating a complementary dual-pathway strategy balancing local refinement with global exploration.

GHG-Net employs a multi-branch architecture to learn multivariate Gaussian distributions of geometric parameters from historical hypothesis data. The network takes as input a set of valid hypotheses from training data and models their distribution in parameter space. GHG-Net consists of an inference block and three parallel prediction branches: one for mean μ, one for variance σ^2, and one for correlation coefficients $r \in [-1, 1]$. The covariance matrix is constructed as $\Sigma_{neural} = SRS + \epsilon I$, where $S = diag(\sigma)$ is the diagonal matrix of standard deviations, R is the correlation matrix constructed from the predicted correlation coefficients, and $\epsilon = 10^{-5}$ ensures positive definiteness and numerical stability. New model parameters are generated via:

$$\theta_{neural} = \mu + L_{neural} z, \tag{7}$$

where L_{neural} is the Cholesky factor of $\Sigma_{neural} = L_{neural} L_{neural}^T$ and $z \sim \mathcal{N}(0, I)$. The geometric models constructed from parameters learned by GHG-Net better conform to actual geometric structures. As shown in Fig. 3, our strategy combines two complementary pathways: a mathematical pathway using Jacobian-based uncertainty for current estimates and a learning pathway using GHG-Net for global exploration based on learned distributions. Pooling hypotheses from both sources enables comprehensive parameter space coverage while maintaining geometric validity.

Mahalanobis Distance-based Clustering

The final stage extracts precise model instances from generated hypotheses through clustering. Prior methods using Euclidean distance fail to capture parameter correlations and anisotropic uncertainties in geometric spaces. We introduce Mahalanobis distance-based clustering that leverages covariance matrices from our Gaussian uncertainty modeling. Compared to Euclidean distance, Mahalanobis distance captures parameter correlations, reflects directional uncertainty through ellipsoidal clustering boundaries, and adapts to heteroscedastic distributions.

Experiments

We report the mean and standard deviation over ten runs for all metrics. We mark the best results in **bold** and the second best results with underline. Please refer to the extended version for implementation details and additional experimental results and discussions.

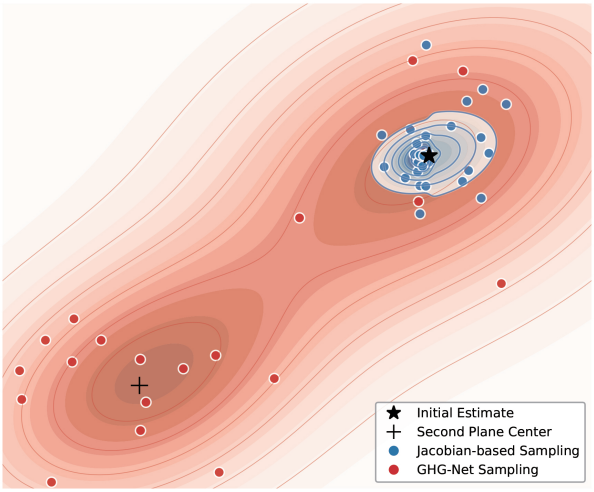

Figure 3: Visualization of parameter space: The blue region represents the Gaussian uncertainty distribution with asymmetric sampling density reflecting inlier distribution, while the red regions show the global distribution learned by GHG-Net. Jacobian-based sampling concentrates in high-inlier-density regions of parameter space, while GHG-Net global sampling explores broader areas, enabling effective discovery of multiple model instances.

Vanishing Point Estimation

We conducted experiments on four datasets: the SU3 and YUD datasets representing Manhattan world scenario, and the NYU-VP and YUD+ datasets containing non-Manhattan world scenario. We compare our method against the robust multi-model fitting methods Progressive-X (Barath and Matas 2019), J-Linkage (Toldo and Fusiello 2008), T-Linkage (Magri and Fusiello 2014), CONSAC (Kluger et al. 2020b), PARSAC (Kluger and Rosenhahn 2024), and the task-specific VP estimators Contrario-VP (Simon, Fond, and Berger 2018), NeurVPS (Zhou et al. 2019), and DeepVP (Lin et al. 2022b). Higher average AUC values indicate superior performance.

Manhattan World. Tab. 1 presents results for the Manhattan scenario. Among task-specific methods using full image information, NeurVPS achieves the best performance on the SU3 dataset. Within the robust estimator category operating on pre-extracted line segments, our method demonstrates superior performance compared to PARSAC. Although NeurVPS maintains the lead on the SU3 dataset, our results are notable since we rely only on pre-extracted line segments rather than full information. Our method outperforms all other robust estimators across all evaluation metrics on the SU3 dataset, with the lowest standard deviations, indicating more stable and reliable results. On the YUD dataset, our method achieves the best performance among

Datasets		SU3 (Zhou et al. 2019)			YUD (Denis, Elder, and Estrada 2008)		
Metrics	GPU/CPU (ms)	AUC @ 1°	AUC @ 3°	AUC @ 5°	AUC @ 3°	AUC @ 5°	AUC @ 10°
task-specific methods (full information)							
Contrario (ECCV'18)	——/767.2	32.56±0.19	67.85±0.17	77.72±0.11	59.86±0.59	72.61±0.47	83.14±0.24
NeurVPS (NIPS'19)	766/——	**79.47**	**92.57**	**95.44**	50.63	63.12	77.58
DeepVP (CVPR'22)	130.3/——	56.63	84.05	90.24	58.19	72.25	84.98
robust estimators (on pre-extracted line segments)							
J-Linkage (ECCV'08)	——/1139	47.66±0.54	74.96±0.49	83.01±0.35	55.75±2.53	68.69±2.22	81.06±1.52
T-Linkage (CVPR'14)	——/271.4	40.57±0.44	71.55±0.26	80.90±0.21	52.19±1.91	66.10±1.41	79.49±0.87
Progressive-X (ICCV'19)	——/28.92	63.14±0.33	80.49±0.40	84.82±0.40	50.41±0.84	60.10±0.76	68.47±0.72
CONSAC (CVPR'20)	3941/2758	51.58±0.18	78.52±0.10	85.69±0.06	59.06±0.48	71.33±0.44	82.79±0.29
PARSAC (AAAI'24)	**4.91/11.59**	67.44±0.14	85.70±0.13	90.17±0.11	63.92±**0.25**	75.90±**0.15**	**86.37**±**0.08**
Ours	6.27/16.83	67.97±**0.11**	85.83±**0.06**	91.22±**0.05**	**64.41**±0.28	**77.28**±0.36	85.23±0.26

Table 1: Vanishing point detection performance and inference time on Manhattan-world (SU3 and YUD datasets)

Datasets		NYU-VP (Kluger et al. 2020b)			YUD+ (Kluger et al. 2020b)		
Metrics	GPU/CPU (ms)	AUC @ 3°	AUC @ 5°	AUC @ 10°	AUC @ 3°	AUC @ 5°	AUC @ 10°
task-specific methods (full information)							
Contrario (ECCV'18)	——/833.2	35.91±0.40	47.61±0.42	61.66±0.38	51.50±0.52	62.81±0.50	72.55±**0.38**
DeepVP (CVPR'22)	176.7/——	**43.54**	**55.87**	**69.53**	48.06	59.57	71.34
robust estimators (on pre-extracted line segments)							
J-Linkage (ECCV'08)	——/1571	30.61±0.81	42.26±0.80	56.58±0.64	48.62±1.29	60.40±1.10	72.36±0.82
T-Linkage (CVPR'14)	——/278.7	30.93±0.64	42.95±0.68	57.75±0.68	46.91±0.66	59.45±0.82	71.74±0.54
Progressive-X (ICCV'19)	——/26.15	38.73±0.17	49.25±0.17	60.71±0.19	50.13±0.78	60.01±0.69	68.53±0.68
CONSAC (CVPR'20)	3843/2841	38.84±0.34	50.58±0.37	64.25±0.38	52.68±0.45	63.73±0.65	74.44±0.76
PARSAC (AAAI'24)	5.16/**9.38**	39.93±0.16	51.65±0.10	64.58±**0.13**	54.98±0.63	65.48±0.37	74.74±0.41
Ours	**4.78**/10.12	40.27±**0.08**	52.93±**0.06**	65.04±0.16	**55.26**±**0.32**	**65.94**±**0.29**	**74.82**±0.42

Table 2: Vanishing point detection performance and inference time on Non-Manhattan (NYU-VP and YUD+ datasets)

robust estimators, and outperforms PARSAC by 0.7% and 1.8% for AUC@3° and AUC@5° respectively. Our method shows a slightly increased inference time on GPU compared to PARSAC, with a modest additional computational overhead, which is reasonable considering the introduction of graph neural network structures.

Non-Manhattan World. Tab. 2 presents results for the non-Manhattan scenario. Among task-specific methods, DeepVP achieves the best performance on NYU-VP via end-to-end learning. Within the robust estimator category, our method demonstrates superior performance across both datasets. On the NYU-VP dataset, our method consistently outperforms PARSAC across all metrics, achieving improvements of 0.85%, 2.48%, and 0.71% for AUC@3°, AUC@5°, and AUC@10° respectively. Notably, our method exhibits exceptional stability with the lowest standard deviations among all robust estimators, particularly for AUC@3° and AUC@5°. On the YUD+ dataset, our method achieves the best performance among robust estimators and demonstrates competitive results compared to task-specific methods, particularly excelling at stricter angular thresholds.

Fundamental Matrix Estimation

We compare our method with Progressive-X (Barath and Matas 2019), Progressive-X+ (Barath et al. 2023), Fast-CP (Ozbay, Camps, and Sznaier 2022), and PARSAC (Kluger and Rosenhahn 2024). CONSAC (Kluger et al. 2020b) has no implementation for F-matrix fitting available. Tab. 3 presents fundamental matrix estimation results on HOPE-F and Adelaide-F datasets. On the HOPE-F dataset, our method achieves the best performance across both metrics, improving ME by 4% and SE by 6% over PARSAC, with superior stability demonstrated by reduced standard deviations. On Adelaide-F, our method outperforms PARSAC by 35% and 27% in ME and SE respectively. For fundamental matrix estimation, our method shows longer inference time on both GPU and CPU compared to PARSAC, a reasonable trade-off for the accuracy gains brought by graph structures and attention mechanism.

Homography Estimation

We compare our method with Progressive-X (Barath and Matas 2019), Progressive-X+ (Barath et al. 2023), CONSAC

Datasets		HOPE-F (Kluger and Rosenhahn 2024)		Adelaide-F (Wong et al. 2011)	
Metrics	GPU/CPU (ms)	ME (%) ↓	SE (pixel) ↓	ME (%) ↓	SE (pixel) ↓
Progressive-X (ICCV'19)	——/1043	22.78±15.64	29.93±105	12.85±14.11	2.19±4.51
Progressive-X+ (CVPR'23)	——/67.38	43.25±15.72	9.71±15.14	6.79±6.64	**0.86±0.89**
Fast-CP (ECCV'22)	——/33.65	24.58±13.32	5.53±6.68	**4.89**±5.26	1.43±1.31
PARSAC (AAAI'24)	**12.25/19.81**	14.97±8.51	3.07±3.39	9.83±4.17	2.80±2.39
Ours	16.36/25.56	**14.36±7.75**	**2.88±3.16**	6.37±**3.26**	2.05±1.94

Table 3: Fundamental matrix estimation on HOPE-F dataset and Adelaide-F dataset

Datasets		SMH (Kluger and Rosenhahn 2024)		Adelaide-H (Wong et al. 2011)	
Metrics	GPU/CPU (ms)	ME (%) ↓	TE (pixel) ↓	ME (%) ↓	TE (pixel) ↓
Progressive-X (ICCV'19)	——/674.2	20.60±15.3	25.5±113	8.05±10.0	**3.14±3.81**
Progressive-X+ (CVPR'23)	——/**501.4**	52.69±18.3	7.19±39.3	7.38±7.64	3.74±5.13
CONSAC (CVPR'20)	4314/58913	33.45±18.5	3.34±25.7	**5.66±7.05**	3.44±7.44
PARSAC (AAAI'24)	**64.0**/1758	20.50±15.5	1.81±20.8	8.63±8.01	5.34±7.36
Ours	76.2/2553	**15.99±14.1**	**0.85±4.05**	7.19±7.27	3.58±4.06

Table 4: Homography estimation on SMH dataset and Adelaide-H dataset

(Kluger et al. 2020b), and PARSAC (Kluger and Rosenhahn 2024). Tab. 4 presents homography estimation results on the challenging SMH and Adelaide-H datasets. On the SMH dataset, our method achieves the best performance across both metrics, with ME of 15.99 (22% improvement over PARSAC) and TE of 0.85 (53% improvement over PARSAC). Notably, our method also demonstrates significantly improved stability with much lower standard deviations. On the Adelaide-H dataset, CONSAC achieves the lowest ME of 5.66, while our method ranks second with 7.19, still outperforming PARSAC by 17%. For the TE metric, our method achieves 3.58, representing a 33% improvement over PARSAC with notably reduced variance. For homography estimation, our method is slower than PARSAC on GPU with a similar difference on CPU, yet remains significantly faster than traditional methods and CONSAC, maintaining the computational efficiency required for real-time applications.

Ablation Experiments

Tab. 5 reveals that each proposed component contributes meaningfully to the overall performance through distinct mechanisms. Comparing our graph neural network architecture with PARSAC's CNN-based method reveals the improvement in performance, reducing ME by 8.7% and TE by 18.2%. This validates our core hypothesis that explicit spatial relationship modeling through dynamic graphs significantly outperforms independent point-wise processing. Incorporating Gaussian uncertainty modeling further reduces ME by 16.9% and TE by 42%, demonstrating that principled hypothesis generation through analytical Jacobian analysis effectively expands the solution space. The complete system achieves the best performance with a 22% improvement in ME and a 53% improvement in TE over the baseline, while also showing dramatically improved stability with standard deviations reduced by up to 80.5%. Notably, the progressive improvements are not merely additive but exhibit synergistic effects. The combination of spatial relationship modeling and uncertainty-aware hypothesis generation creates a multiplicative benefit, where enhanced feature representations enable more effective utilization of the expanded hypothesis space. Comprehensive ablation experiments are presented in the extended version.

Configuration	ME (%) ↓	TE (pixel) ↓
PARSAC (CNN Baseline)	20.5±15.5	1.81±20.8
Ours (GNN)	18.7±15.3	1.48±12.3
Ours (GNN) + Gauss (only)	17.0±14.6	1.05±6.86
Ours (GNN) + GHG-Net (only)	17.4±14.2	1.14±5.67
Ours (GNN) + Gauss + GHG-Net	**15.9±14.1**	**0.85±4.05**

Table 5: Progressive component analysis on SMH dataset

Conclusion

In this paper, we propose a robust multi-model fitting method that integrates Gaussian uncertainty modeling with graph neural networks. Comprehensive experiments on vanishing point detection, fundamental matrix estimation, and homography estimation demonstrate that our method consistently outperforms existing methods in both accuracy and robustness while maintaining computational efficiency. In future work, we will explore more efficient graph construction strategies for real-time applications. Additionally, we plan to extend our method to other geometric vision tasks and explore real-world applications in autonomous driving and 3D reconstruction where robust multi-model fitting is essential for safety-critical decision making.

Acknowledgments

This work was supported in part by the Natural Science Foundation of Fujian Province of China under Grant 2025J01248, and in part by the Science and Technology Program of Quanzhou under Grant 2024QZC001R, and in part by the Science and Technology Program of Fujian Province under Grants 2024T3055 and 2025T3027.

References

Abu Alhaija, H.; Mustikovela, S. K.; Mescheder, L.; Geiger, A.; and Rother, C. 2018. Augmented reality meets computer vision: Efficient data generation for urban driving scenes. *International Journal of Computer Vision*, 126: 961–972.

Antunes, M.; and Barreto, J. P. 2013. A global approach for the detection of vanishing points and mutually orthogonal vanishing directions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1336–1343.

Barath, D.; and Matas, J. 2018. Multi-class model fitting by energy minimization and mode-seeking. In *Proceedings of the European Conference on Computer Vision*, 221–236.

Barath, D.; and Matas, J. 2019. Progressive-X: Efficient, anytime, multi-model fitting algorithm. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 3780–3788.

Barath, D.; Rozumnyi, D.; Eichhardt, I.; Hajder, L.; and Matas, J. 2023. Finding geometric models by clustering in the consensus space. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 5414–5424.

Barinova, O.; Lempitsky, V.; Tretiak, E.; and Kohli, P. 2010. Geometric image parsing in man-made environments. In *Proceedings of the European Conference on Computer Vision*, 57–70.

Brachmann, E.; and Rother, C. 2018. Learning less is more-6D camera localization via 3D surface regression. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4654–4662.

Brachmann, E.; and Rother, C. 2019. Neural-Guided RANSAC: Learning where to sample model hypotheses. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*.

Cavalli, L.; Barath, D.; Pollefeys, M.; and Larsson, V. 2023. Consensus-Adaptive RANSAC. arXiv:2307.14030.

Cavalli, L.; Pollefeys, M.; and Barath, D. 2022. NeFSAC: Neurally filtered minimal samples. In *Proceedings of the European Conference on Computer Vision*, 351–366.

Denis, P.; Elder, J. H.; and Estrada, F. J. 2008. Efficient edge-based methods for estimating manhattan frames in urban imagery. In *Proceedings of the European Conference on Computer Vision*, 197–210.

Farina, M.; Magri, L.; Menapace, W.; Ricci, E.; Golyanik, V.; and Arrigoni, F. 2023. Quantum multi-model fitting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 13640–13649.

Fischler, M. A.; and Bolles, R. C. 1981. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6): 381–395.

Guo, H.; Lu, Y.; Xiao, G.; Lin, S.; and Wang, H. 2022. Triplet relationship guided sampling consensus for robust model estimation. *IEEE Signal Processing Letters*, 29: 817–821.

Guo, H.; Xiao, G.; Su, L.; Li, T.; Wang, D.-H.; and Wang, H. 2024. Second-order proximity guided sampling consensus for robust model fitting. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(11): 10704–10717.

Heinly, J.; Schonberger, J. L.; Dunn, E.; and Frahm, J.-M. 2015. Reconstructing the world* in six days*(As captured by the Yahoo 100 million image dataset). In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3287–3295.

Isack, H.; and Boykov, Y. 2012. Energy-based geometric multi-model fitting. *International Journal of Computer Vision*, 97(2): 123–147.

Kluger, F.; Ackermann, H.; Yang, M. Y.; and Rosenhahn, B. 2020a. Temporally consistent horizon lines. In *2020 IEEE International Conference on Robotics and Automation*, 3161–3167.

Kluger, F.; Brachmann, E.; Ackermann, H.; Rother, C.; Yang, M. Y.; and Rosenhahn, B. 2020b. CONSAC: Robust multi-model fitting by conditional sample consensus. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4634–4643.

Kluger, F.; and Rosenhahn, B. 2024. PARSAC: Accelerating robust multi-model fitting with parallel sample consensus. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

Lezama, J.; Grompone von Gioi, R.; Randall, G.; and Morel, J.-M. 2014. Finding vanishing points via point alignments in image primal and dual domains. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 509–515.

Lin, S.; Luo, H.; Yan, Y.; Xiao, G.; and Wang, H. 2022a. Co-clustering on bipartite graphs for robust model fitting. *IEEE Transactions on Image Processing*, 31: 6605–6620.

Lin, S.; Xiao, G.; Yan, Y.; Suter, D.; and Wang, H. 2019. Hypergraph optimization for multi-structural geometric model fitting. In *Proceedings of the AAAI conference on Artificial Intelligence*, 8730–8737.

Lin, Y.; Wiersma, R.; Pintea, S. L.; Hildebrandt, K.; Eisemann, E.; and Van Gemert, J. C. 2022b. Deep vanishing point detection: Geometric priors make dataset variations vanish. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 6103–6113.

Liu, S.; Zhou, Y.; and Zhao, Y. 2021. VaPiD: A rapid vanishing point detector via learned optimizers. In *Proceedings of the European Conference on Computer Vision*, 12839–12848.

Lowe, D. G. 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60: 91–110.

Magri, L.; and Fusiello, A. 2014. T-linkage: A continuous relaxation of j-linkage for multi-model fitting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3954–3961.

Magri, L.; Fusiello, A.; et al. 2015. Robust multiple model fitting with preference analysis and low-rank approximation. In *Procedings of the British Machine Vision Conference 2015*, 20–1.

Mur-Artal, R.; and Tardós, J. D. 2017. ORB-SLAM2: An open-source SLAM system for monocular, stereo, and RGB-D cameras. *IEEE Transactions on Robotics*, 33(5): 1255–1262.

Ozbay, B.; Camps, O.; and Sznaier, M. 2022. Fast two-view motion segmentation using christoffel polynomials. In *Proceedings of the European Conference on Computer Vision*, 1–19.

Pautrat, R.; Barath, D.; Larsson, V.; Oswald, M. R.; and Pollefeys, M. 2023. Deeplsd: Line segment detection and refinement with deep image gradients. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 17327–17336.

Purkait, P.; Chin, T.-J.; Sadri, A.; and Suter, D. 2016. Clustering with hypergraphs: The case for large hyperedges. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(9): 1697–1711.

Sattler, T.; Leibe, B.; and Kobbelt, L. 2016. Efficient & effective prioritized matching for large-scale image-based localization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(9): 1744–1756.

Schonberger, J. L.; and Frahm, J.-M. 2016. Structure-from-motion revisited. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4104–4113.

Simon, G.; Fond, A.; and Berger, M.-O. 2018. A-contrario horizon-first vanishing point detection using second-order grouping laws. In *Proceedings of the European Conference on Computer Vision*, 318–333.

Tardif, J.-P. 2009. Non-iterative approach for fast and accurate vanishing point detection. In *Proceedings of the European Conference on Computer Vision*, 1250–1257.

Toldo, R.; and Fusiello, A. 2008. Robust multiple structures estimation with j-linkage. In *Proceedings of the European Conference on Computer Vision*, 537–547.

Vincent, E.; and Laganiere, R. 2001. Detecting planar homographies in an image pair. In *ISPA 2001. Proceedings of the 2nd International Symposium on Image and Signal Processing and Analysis. In conjunction with 23rd International Conference on Information Technology Interfaces*, 182–187.

Von Gioi, R. G.; Jakubowicz, J.; Morel, J.-M.; and Randall, G. 2008. LSD: A fast line segment detector with a false detection control. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(4): 722–732.

Wandt, B.; and Rosenhahn, B. 2019. Repnet: Weakly supervised training of an adversarial reprojection network for 3D human pose estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7782–7791.

Wang, H.; Xiao, G.; Yan, Y.; and Suter, D. 2018. Searching for representative modes on hypergraphs for robust geometric model fitting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(3): 697–711.

Wei, T.; Matas, J.; and Barath, D. 2023. Adaptive reordering sampler with neurally guided MAGSAC. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 18163–18173.

Wei, T.; Patel, Y.; Shekhovtsov, A.; Matas, J.; and Barath, D. 2023. Generalized differentiable RANSAC. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 17649–17660.

Wildenauer, H.; and Hanbury, A. 2012. Robust camera self-calibration from monocular images of Manhattan worlds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2831–2838.

Wong, H. S.; Chin, T.-J.; Yu, J.; and Suter, D. 2011. Dynamic and hierarchical multi-structure geometric model fitting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 1044–1051.

Wu, J.; Zhang, L.; Liu, Y.; and Chen, K. 2021. Real-time vanishing point detector integrating under-parameterized RANSAC and hough transform. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 3712–3721.

Xiao, G.; Wang, H.; Yan, Y.; and Zhang, L. 2016. Mode seeking on graphs for geometric model fitting via preference analysis. *Pattern Recognition Letters*, 83: 294–302.

Xiao, G.; Yu, J.; Ma, J.; Fan, D.-P.; and Shao, L. 2024. Latent semantic consensus for deterministic geometric model fitting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(9): 6139–6153.

Zhai, M.; Workman, S.; and Jacobs, N. 2016. Detecting vanishing points using global image context in a Non-manhattan world. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5657–5665.

Zhou, Y.; Qi, H.; Huang, J.; and Ma, Y. 2019. Neurvps: Neural vanishing point scanning via conic convolution. In *Advances in Neural Information Processing Systems*.