BA-GS: Bayesian Adaptive Gaussian Splatting for SFM-Free 3D Reconstruction

Supplementary Material

Justification of Gaussian Approximation

This section provides a theoretical proof for treating both prediction and observation noise in the BA-GS pipeline as Gaussian random variables, and for employing Kalman or Extended-Kalman style updates under this assumption.

Assumptions.

We make the following mild and standard assumptions, which are explicitly stated here for clarity:

(1) **(Multi-source small perturbations)** The observation noise v can be decomposed as the sum of many small, independent noise composition:

$$v = \sum_{i=1}^{n} \delta_i \tag{12}$$

where each δ_i has finite mean $\mathbb{E}[\delta_i]$ and finite variance $\mathrm{Var}(\delta_i)$.

(2) **(Finite aggregated variance)** The total variance $\sigma_v^2 = \sum_{i=1}^{n} \mathrm{Var}(\delta_i)$ is finite and nonzero.

(3) **(Local smoothness)** The forward function $h = \mathbb{R}^d \to \mathbb{R}^m$ (e.g., 3D to 2D projection + compositing) is locally smooth so that a first-order Taylor expansion around a working estimate $\hat{x}$ provides a good approximation for states in a neighborhood of $\hat{x}$.

Aggregated noise and approximation

Under assumption (1) and (2), the observation noise v results from the superposition of many small disturbances. By the CLT, as the number of contributing factors grows,

$$\frac{v - \mathbb{E}[v]}{\sqrt{\mathrm{Var}(v)}} \xrightarrow{d} \mathcal{N}(0, 1). \tag{13}$$

Even the individual sources of noise are not Gaussian, their sum tends toward a gaussian distribution. Hence, we approximate

$$v \approx \mathcal{N}(\mu_v, \Sigma_v), \tag{14}$$

with mean and covariance

$$\mu_v = \sum_i \mathbb{E}[\delta_i], \qquad \Sigma_v = \sum_i \mathrm{Var}(\delta_i). \tag{15}$$

where the δ_i indicates the components of aggregated noise.

L_2 residuals and Gaussian likelihood. Consider the observation model

$$z = h(x) + v \tag{16}$$

If $v \sim \mathcal{N}(0, R)$, the negative log-likelihood is

$$-\log p(z \mid x) = \tfrac{1}{2}(z - h(x))^\top R^{-1}(z - h(x)). \tag{17}$$

Thus minimizing an L_2 or weighted L_2 error is equivalent to maximizing a Gaussian likelihood. In other words, whenever L_2 reconstruction losses are used, the underlying statistical model implicitly assumes Gaussian observation noise. In practice, however, standard 3DGS often employs an L_1 or SSIM loss, which instead implies a Laplacian observation model. Since exact Bayesian inference under this Laplacian model is intractable and would require expensive MCMC sampling, we employ a Gaussian approximation to enable analytical, closed-form updates. Individual pixel-wise residuals follow a Laplacian (L_1) distribution, the actual parameter update step aggregates gradients from many pixels. By the CLT, the noise distribution of these aggregated gradients converges to a Gaussian, further validating the use of the KF for position denoising.

Local Linearization and EKF Updates

In the 3DGS framework, the observation function h is highly nonlinear, rendering the standard linear Kalman Filter inapplicable. To address this, we employ the Extended Kalman Filter (EKF) by locally linearizing h around the current state estimate $\hat{x}$ using a first-order Taylor expansion:

$$h(x) \approx h(\hat{x}) + H(\hat{x})(x - \hat{x}), \tag{18}$$

where $H(\hat{x}) = \frac{\partial h}{\partial x}\big|_{x=\hat{x}}$ is the Jacobian matrix evaluated at $\hat{x}$. Substituting this into the standard observation model $z = h(x) + v$ yields the linearized model:

$$z \approx h(\hat{x}) + H(\hat{x})(x - \hat{x}) + \eta, \tag{19}$$

where $\eta \sim \mathcal{N}(0, R)$ encapsulates both the additive Gaussian measurement noise and the first-order linearization residuals.

Assuming a Gaussian prior for the state $x \sim \mathcal{N}(\hat{x}, P)$, we define the measurement residual as $\tilde{z} \triangleq z - h(\hat{x})$. This allows us to formulate the approximate linear Gaussian relation:

$$\tilde{z} \approx H(\hat{x})(x - \hat{x}) + \eta. \tag{20}$$

Under this locally linearized model, the EKF updates share the same algebraic structure as the standard Kalman Filter, with the Jacobian H replacing the linear observation matrix. This justifies our use of standard KF notation in the main text. The posterior mean and covariance are updated as follows:

$$K = PH^\top (HPH^\top + R)^{-1}, \tag{21}$$

$$\hat{x}^+ = \hat{x} + K(z - h(\hat{x})), \tag{22}$$

$$P^+ = (I - KH)P. \tag{23}$$

Here, $\hat{x}$ and P denote the prior state estimate and its corresponding error covariance matrix respectively. $z - h(\hat{x})$ represents the measurement residual which stands for capturing the discrepancy between the actual observation z and the predicted measurement $h(\hat{x})$. The gain K optimally weights this residual by balancing the prior uncertainty P against the measurement noise covariance R. Besides, $\hat{x}^+$ and P^+ represent the updated posterior state estimate and its reduced covariance matrix after incorporating the new observation, with I being the identity matrix.

The validity of this approximation depends on the local smoothness of h (Assumption 3) and on the aggregated noise being approximately normal (Assumptions 1 and 2). When these conditions hold, the EKF provides a computationally efficient extension of the Kalman framework for nonlinear systems.

However, scaling this extended framework to 3D scenes with millions of Gaussian primitives introduces new computational bottlenecks, particularly in maintaining the covariance matrices. While we acknowledge the spatial correlation of gradients, explicitly modeling the full covariance is computationally expensive. Thus, we adopt a diagonal approximation, which is a standard practice in bundle adjustment. Moreover, BA-GS relies on the iterative nature of the optimization to resolve these correlations over iterations.

Synthesis. We proved that:

- Multiple small, heterogeneous noise sources obeys Gaussian distribution through the CLT.
- The L_2 loss corresponds to a Gaussian likelihood.
- Under local smoothness, linearization preserves the Gaussian form and enables closed-form EKF updates.

Discussion on Failure Cases

Despite the robustness of the Gaussian approximation and filtering mechanism, there are specific challenging scenarios where the underlying assumptions may break down:

- **Specular surfaces:** Highly view-dependent reflections violate the multi-view geometric consistency inherent in our observation model. Instead of zero-mean noise, this results in consistent geometric drift during optimization.
- **Textureless regions:** In areas lacking texture, photometric gradients approach zero. This lack of gradient information leads to unobservable states where the Kalman gain effectively ignores observations, relying solely on prior estimates without meaningful corrections.

More Results

We provide extensive visual comparisons on LLFF, MVImgNet, and Tanks&Temples. Unless otherwise specified, all visualizations follow the same rendering setup as the main paper. We highlight reflective regions, such as the TV reflections in the LLFF room scene and the piano reflections in the Tanks&Temples Church scene.

In Fig. 6, the effect of VB-GMM on initialization is evident that the primitive distribution becomes more structured and less redundant, providing a stronger basis for subsequent optimization.

Figure 6. Comparison of point clouds on three datasets: LLFF, MVImgNet, and Tanks and Temples (rows). Each row shows results from InstantSplat (left), our BA-GS method (middle), and their overlay: InstantSplat (Red), BA-GS (Green).

Figure 7. Zoom-in comparison(Tanks, 18 views setting). Leftmost column: Ground Truth(full / cropped). Second and third columns: DropGaussian (full / cropped). Fourth and fifth columns: InstantSplat (full / cropped).

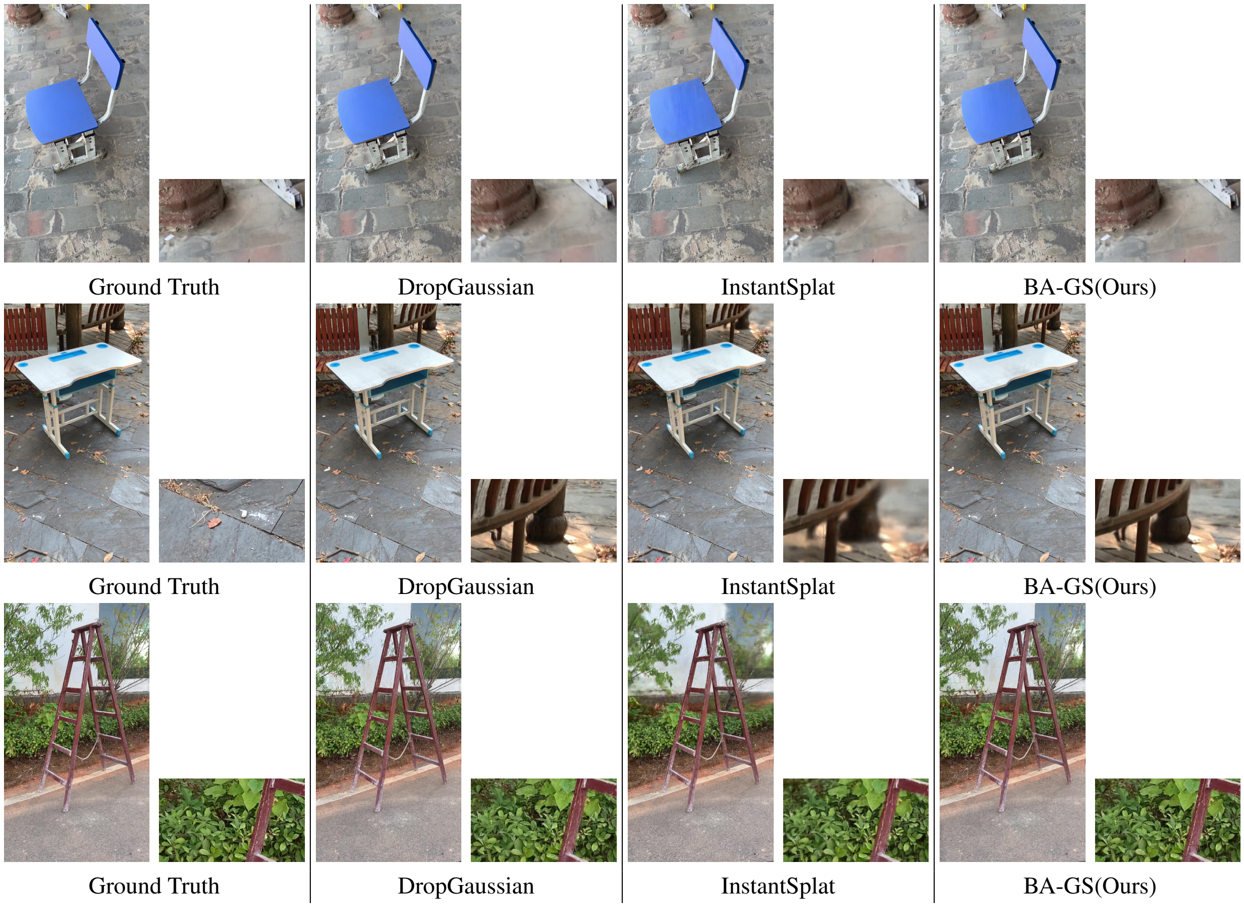

Figure 8. Zoom-in comparison(MVimgNet, 18 views setting). Leftmost column: Ground Truth(full / cropped). Second and third columns: DropGaussian (full / cropped). Fourth and fifth columns: InstantSplat (full / cropped).

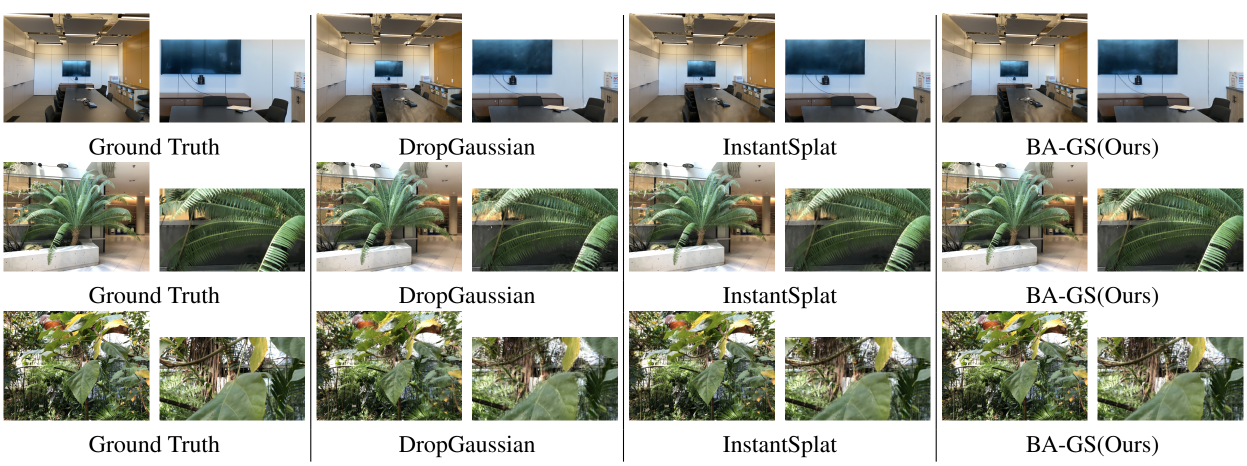

Figure 9. Zoom-in comparison(LLFF, 18 views setting). Leftmost column: Ground Truth(full / cropped). Second and third columns: DropGaussian (full / cropped). Fourth and fifth columns: InstantSplat (full / cropped).