SUPPLEMENTARY MATERIALS

Due to page limitations, we provide additional detailed explanations related to the paper below.

A. NOMENCLATURE

σ	**noise standard deviations**
$\mathrm{SIM}(3)$	**Special Similitude Group in 3D**
$\mathfrak{sim}(3)$	**Lie algebra associated with** $\mathrm{SIM}(3)$
$S = \{R, t, s\}$	**Transformation parameters including rotation, translation, and scale**
$\xi = (\omega, \mathbf{v}, \sigma)$	**Twist vector in the Lie algebra**
$\mathrm{SE}(3)$	**Special Euclidean Group in 3D**
$\mathrm{SO}(3)$	**Special Orthogonal Group in 3D**
$T = \{R, t\}$	**Homogeneous transformation matrix with rotation and translation**
$Ad(T)$	**Adjoint matrix of** T
$\wedge$	**convert vector to skew-symmetric matrix**
$\vee$	**convert skew-symmetric matrix to vector**

B. The Derivation of Jacobian

The following outlines the propagation process from multiple sources of uncertainty to joint uncertainty. We calculate the Jacobian to estimate uncertainty, focusing exclusively on rotation and translation. The scale factor σ in $\xi = (\omega, \mathbf{v}, \sigma)$ is not considered. We derive the partial derivatives of $h(\xi_j^\beta)$ with respect to ξ_j^β

$$J_{\xi_j^\beta} = \frac{\partial h}{\partial \xi_j^\beta} = \frac{h(\xi_j^\beta + \Delta\xi_j^\beta) - h(\xi_j^\beta)}{\Delta\xi_j^\beta} \tag{1}$$

According to the Baker-Campbell-Hausdorff (BCH) formula, when $\Delta\xi$ is small relative to ξ, the following linear relationship holds

$$\exp\left(\Delta\xi^\wedge\right)\exp\left(\xi^\wedge\right) \approx \exp\left(\left(J_l^{-1}\Delta\xi + \xi\right)^\wedge\right) \tag{2}$$

By substituting (**??**) into the cost function, the following expression is obtained

$$\left(T_i^\alpha T_{i,j}^{\alpha,\beta}\right)^{-1} T_\beta^\alpha \exp\left(\left(\xi_j^\beta + \Delta\xi_j^\beta\right)^\wedge\right) \approx \left(T_i^\alpha T_{i,j}^{\alpha,\beta}\right)^{-1} T_\beta^\alpha \exp\left(\left(\mathcal{J}_l\left(\xi_j^\beta\right)\Delta\xi_j^\beta\right)^\wedge\right)\exp(\xi_j^{\beta\wedge}) \tag{3}$$

Where

$$\mathcal{J}_l\left(\xi_j^\beta\right) \approx I + \frac{1}{2}\begin{bmatrix} \phi_{\xi_j^\beta}^\wedge & \rho_{\xi_j^\beta}^\wedge \\ 0 & \phi_{\xi_j^\beta}^\wedge \end{bmatrix} \tag{4}$$

The left Jacobian $\mathcal{J}_l(\xi_j^\beta)$ represents the first-order approximation of the perturbation. According to

$$\exp\left(\xi^\wedge\right)T = T\exp\left(\left(\mathrm{Ad}(T^{-1})\xi\right)^\wedge\right) \tag{5}$$

Hence, integrating (**??**) with (**??**) produces the following result

$$h(\xi_j^\beta + \Delta\xi_j^\beta) = \left(T_i^a T_{i,j}^{\alpha,\beta}\right)^{-1} T_\beta^\alpha T_j^\beta \exp\left(\mathrm{Ad}\left((T_j^\beta)^{-1}\right)\mathcal{J}_l(\xi_j^\beta)\Delta\xi_j^\beta\right) \tag{6}$$

Given that

$$\mathrm{In}\left(\exp(\xi^\wedge)\exp(\Delta\xi^\wedge)\right)^\vee \approx \mathcal{J}_r^{-1}\Delta\xi + \xi \tag{7}$$

The Jacobian matrix $J_{\xi_j^\beta}$ from (**??**) is given by

$$\begin{aligned} J_{\xi_j^\beta} &\approx \frac{\mathcal{J}_r\left(h(\xi_j^\beta)\right)^{-1}\mathrm{Ad}\left((T_j^\beta)^{-1}\right)\mathcal{J}_l(\xi_j^\beta)\Delta\xi_j^\beta + h(\xi_j^\beta) - h(\xi_j^\beta)}{\Delta\xi_j^\beta} \\ &= \mathcal{J}_r\left(h(\xi_j^\beta)\right)^{-1}\mathrm{Ad}\left((T_j^\beta)^{-1}\right)\mathcal{J}_l(\xi_j^\beta) \end{aligned} \tag{8}$$

Here, the right Jacobian $\mathcal{J}_r$ with respect to $h(\xi_j^\beta)$ is approximated as follows

$$\mathcal{J}_r^{-1}(h) \approx I + \frac{1}{2}\begin{bmatrix} \phi_h^\wedge & \rho_h^\wedge \\ 0 & \phi_h^\wedge \end{bmatrix} \tag{9}$$

Similarly, we derive the Jacobian matrix for ξ_i^a based on the process described above.

$$\begin{aligned} J_{\xi_i^a} &= \frac{\partial h}{\partial \xi_i^a} = \frac{h(\xi_i^a + \Delta\xi_i^a) - h(\xi_i^a)}{\Delta\xi_i^a} \\ &= \frac{In\left(\left(\exp\left((\xi_i^a + \Delta\xi_i^a)^\wedge\right) T_{i,j}^{\alpha,\beta}\right)^{-1} T_\beta^\alpha T_j^\beta\right)^\vee - In\left(\left(\exp\left(\xi_i^{\alpha\wedge}\right) T_{i,j}^{\alpha,\beta}\right)^{-1} T_\beta^\alpha T_j^\beta\right)^\vee}{\Delta\xi_i^\alpha} \\ &\approx \frac{-\mathcal{J}_r\left(h\left(\xi_i^\alpha\right)\right)^{-1} Ad\left(\left(T_\beta^\alpha T_j^\beta\right)^{-1}\right) \mathcal{J}_l(\xi_i^\alpha)\Delta\xi_i^\alpha + h\left(\xi_j^\beta\right) - h\left(\xi_j^\beta\right)}{\Delta\xi_i^\alpha} \\ &= -\mathcal{J}_r\left(h\left(\xi_i^a\right)\right)^{-1} Ad\left(\left(T_\beta^\alpha T_j^\beta\right)^{-1}\right) \mathcal{J}_l(\xi_i^a) \end{aligned} \tag{10}$$

Then, we derive the Jacobian matrix for $\xi_{i,j}^{\alpha,\beta}$

$$\begin{aligned} J_{\xi_{i,j}^{\alpha,\beta}} &= \frac{\partial h}{\partial \xi_{i,j}^{\alpha,\beta}} = \frac{h(\xi_{i,j}^{\alpha,\beta} + \Delta\xi_{i,j}^{\alpha,\beta}) - h(\xi_{i,j}^{\alpha,\beta})}{\Delta\xi_{i,j}^{\alpha,\beta}} \\ &= \frac{In\left(\left(T_i^\alpha \exp\left((\xi_{i,j}^{\alpha,\beta} + \Delta\xi_{i,j}^{\alpha,\beta})^\wedge\right)\right)^{-1} T_\beta^\alpha T_j^\beta\right)^\vee - In\left(T_i^\alpha\left(\exp\left(\xi_{i,j}^{\alpha,\beta\wedge}\right)\right)^{-1} T_\beta^\alpha T_j^\beta\right)^\vee}{\Delta\xi_{i,j}^{\alpha,\beta}} \\ &\approx \frac{-\mathcal{J}_r\left(h\left(\xi_{i,j}^{\alpha,\beta}\right)\right)^{-1} Ad\left(\left((T_i^\alpha)^{-1} T_\beta^\alpha T_j^\beta\right)^{-1}\right) \mathcal{J}_l(\xi_{i,j}^{\alpha,\beta})\Delta\xi_{i,j}^{\alpha,\beta} + h\left(\xi_j^\beta\right) - h\left(\xi_j^\beta\right)}{\Delta\xi_{i,j}^{\alpha,\beta}} \\ &= -\mathcal{J}_r\left(h\left(\xi_{i,j}^{\alpha,\beta}\right)\right)^{-1} Ad\left(\left((T_i^\alpha)^{-1} T_\beta^\alpha T_j^\beta\right)^{-1}\right) \mathcal{J}_l(\xi_{i,j}^{\alpha,\beta}) \end{aligned} \tag{11}$$

Thus far, we have completed the derivation.

C. Adaptive σ-consensus Approach

Kernel Density Estimation (KDE) is powerful a non-parametric technique for estimating probability density functions. Unlike parametric methods, KDE does not rely on prior assumptions about the underlying distribution, such as normal or exponential distributions, offering greater flexibility in modeling diverse datasets.

To evaluate the distribution of residuals, we denote the residual as r, with residuals computed as the Euclidean distance from the estimated model. The density function of the residuals for inliers is defined as

$$g(r) = \frac{1}{n}\sum_{i=1}^{n} K_h(r - r_i) \tag{12}$$

where n is the number of residual samples, h is the bandwidth parameter controlling the smoothness of the density estimate, and K is the kernel function. The term accounts for the contribution of each sample r_i to the density at the point r, with bandwidth adjustment. Specifically, $K_h(r)$ is expressed as

$$K_h(r) = \frac{1}{h}K\left(\frac{r}{h}\right) \tag{13}$$

where the kernel function K determines the shape of the weighting function applied to residuals. For this work, we adopt the Gaussian kernel, a widely used choice due to its smooth and differentiable properties. The Gaussian kernel is defined as

$$K(r) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{r^2}{2}\right) \tag{14}$$

When incorporating the bandwidth parameter h, the Gaussian kernel is reformulated as

$$K_h(r - r_i) = \frac{1}{h\sqrt{2\pi}}\exp\left(-\frac{(r - r_i)^2}{2h^2}\right) \tag{15}$$

Substituting (**??**) into (**??**) yields the explicit expression for the residual density estimate

$$g(r) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h\sqrt{2\pi}}\exp\left(-\frac{(r-r_i)^2}{2h^2}\right) \tag{16}$$

In this formulation, the parameter h plays a critical role. KDE identifies low-density regions in the data by estimating the density distribution, aiding in the detection of outliers. The threshold for distinguishing inliers and outliers is closely related to the bandwidth h, as h directly affects the smoothing degree of the Kernel Density Estimation (KDE), thereby indirectly altering the distribution characteristics of density values. If the bandwidth is too large, the density curve becomes overly smooth, causing the density values of all points to converge. Conversely, if the bandwidth is too small, the density curve becomes too sharp, with local high-density peaks forming around each data point. Given the significant impact of bandwidth on the threshold for inlier and outlier classification, selecting an appropriate bandwidth is critical.

Given the use of a Gaussian kernel for smoothing, the bandwidth h is set according to Silverman's Rule of Thumb, expressed as

$$h = k\sigma n^{-1/5} \tag{17}$$

where σ denotes the standard deviation of the data, $k = 1.06$ is a constant, and n is the number of data points. This formulation ensures that h dynamically adjusts according to the data distribution, allowing for effective density estimation that adapts to the scale of σ. Substituting (**??**) into (**??**) yields the density function conditioned on σ

$$g(r \mid \sigma) = \frac{1}{n}\sum_{i=1}^{n}\frac{n^{1/5}}{k\sigma\sqrt{2\pi}}\exp\left(-\frac{n^{2/5}(r-r_i)^2}{2k^2\sigma^2}\right) \tag{18}$$

This formulation provides a probabilistic framework to assess the likelihood of a point being an inlier based on its residual density. The likelihood estimate is used as a weight by marginalizing over the uncertainty parameter σ. This process is defined as

$$w(r) = \int g(r \mid \sigma) f(\sigma)\,\mathrm{d}\sigma \tag{19}$$

where $f(\sigma)$ is the probability density function of σ. Assuming σ is uniformly distributed as $\sigma \sim \mathcal{U}(0, \sigma_{max})$, the weight w(r) can be expressed as

$$w(r) = \frac{1}{\sigma_{\max}}\int_0^{\sigma_{\max}}\frac{1}{k\sqrt{2\pi}\sigma n^{4/5}}\sum_{i=1}^{n}\exp\left(-\frac{n^{2/5}(r-r_i)^2}{2k^2\sigma^2}\right)\mathrm{d}\sigma \tag{20}$$

Given the independence of each sample r_i in kernel density estimation, we first marginalize a single kernel function and then sum the resulting values. For each sample r_i, the marginalized density function is given by

$$f(r \mid \sigma) = \frac{n^{1/5}}{k\sigma\sqrt{2\pi}}\exp\left(-\frac{n^{2/5}(r-r_i)^2}{2k^2\sigma^2}\right) \tag{21}$$

By substituting the uniform distribution $\frac{1}{\sigma_{max}}$ into the integral, we obtain

$$w(r) = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\sigma_{max}} f(r \mid \sigma)\frac{1}{\sigma_{max}}\,\mathrm{d}\sigma = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\sigma_{max}}\frac{n^{1/5}}{k\sigma\sqrt{2\pi}}\exp\left(-\frac{n^{2/5}(r-r_i)^2}{2k^2\sigma^2}\right)\frac{1}{\sigma_{max}}\,\mathrm{d}\sigma \tag{22}$$

Combining the constant terms simplifies the expression to

$$w(r) = \frac{1}{nk\sigma_{max}\sqrt{2\pi}}\sum_{i=1}^{n} n^{1/5}\int_0^{\sigma_{max}}\frac{1}{\sigma}\exp\left(-\frac{n^{2/5}(r-r_i)^2}{2k^2\sigma^2}\right)\mathrm{d}\sigma \tag{23}$$

To simplify the integral in the kernel density estimation weight computation, we perform a variable substitution. Let

$$u = \frac{1}{\sigma^2}, \quad \sigma = \frac{1}{\sqrt{u}}, \quad \frac{1}{\sigma}d\sigma = -\frac{1}{2}\frac{1}{u}du \tag{24}$$

The integration limits change such that when $\sigma = 0$,$u \to \infty$, and when $\sigma = \sigma_{max}$, $u = \frac{1}{\sigma_{max}^2}$. Substituting these transformations into the integral for the marginalized density function, we have

$$f(r \mid \sigma) = \int_{1/\sigma_{max}^2}^{\infty} -\frac{\exp\left(-\frac{n^{2/5}(r-r_i)^2}{2k^2}u\right)}{2u}\,\mathrm{d}u \tag{25}$$

This integral corresponds to a generalized form of the exponential integral, expressed as

$$\int_z^{\infty}\frac{\exp(-\alpha t)}{t}\,\mathrm{d}t = E_1(\alpha z) \tag{26}$$

where $E_1(z)$ is the exponential integral, defined as

$$E_1(z) = \int_z^{\infty} \frac{\exp(-t)}{t} \, \mathrm{d}t \tag{27}$$

By matching parameters in the integral, we identify the following

$$\alpha = \frac{n^{2/5}(r - r_i)^2}{2k^2}, \quad z = \frac{1}{\sigma_{max}^2} \tag{28}$$

Thus, the integral evaluates to

$$\int_{1/\max^2}^{\infty} \frac{\exp\left(-\frac{n^{2/5}(r-r_i)^2}{2k^2} u\right)}{2u} \, \mathrm{d}u = \frac{1}{2} E_1 \left(\frac{n^{2/5}(r - r_i)^2}{2k^2 \sigma_{max}^2}\right) \tag{29}$$

Substituting this result into the KDE weight formula, the weight function $w(r)$ is expressed as

$$w\left(r\right) = \frac{1}{n} \sum_{i=1}^{n} \frac{n^{1/5}}{2k\sqrt{2\pi}\sigma_{\max}} E_1 \left(\frac{n^{2/5}\left(r - r_i\right)^2}{2k^2 \sigma_{\max}^2}\right) \tag{30}$$

This formulation offers a efficient approach to incorporating residual into kernel density-based weighting.