LVBA: LiDAR-Visual Bundle Adjustment for RGB Point Cloud Mapping (Supplementary)

Please note that equation numbers and section numbers from the main manuscript are labeled in this letter in red.

Appendix A
Covariance of Simplified Photometric Camera Model

Our method follows [1] and makes some simplification with it. When a ray is projected to a pixel on the camera, the photometric camera model can be represented as

$$\mathbf{I}(\boldsymbol{\rho}) = \mathbf{f}(\tau V(\boldsymbol{\rho})\boldsymbol{\gamma}) \tag{1}$$

where $\boldsymbol{\rho}$ is the pixel position, $\boldsymbol{\gamma}$ is the radiance of the ray, τ is the camera exposure time, $\mathbf{f}(\cdot)$ is the camera response function (CRF), and $V(\cdot)$ is the vignetting factor accounting for the lens vignetting effect. Since the CRF and vignetting factor of the camera are small but hard to obtain after the sequence is collected, we simply ignore them by setting them to one, leading to a camera photometric model as below

$$\mathbf{I} = \mathbf{f}(\tau V(\boldsymbol{\rho})\boldsymbol{\gamma}) \overset{V(\cdot)=1}{\approx} f(\tau\boldsymbol{\gamma}) \overset{\mathbf{f}(\mathbf{x})=\mathbf{x}}{=} \tau\boldsymbol{\gamma} \tag{2}$$

Further, we model all sources of measurement noise (e.g. noise from AD transition on camera CMOS) as a Gaussian random noise $\boldsymbol{\delta}_c \sim \mathcal{N}(0, \boldsymbol{\Sigma}_c)$, which will lead to our simplified measurement model:

$$\mathbf{I} = \boldsymbol{\gamma}\tau + \boldsymbol{\delta}_c; \boldsymbol{\delta}_c \sim \mathcal{N}(0, \boldsymbol{\Sigma}_c); \boldsymbol{\Sigma}_c = \begin{bmatrix} \sigma_c^2 & 0 & 0 \\ 0 & \sigma_c^2 & 0 \\ 0 & 0 & \sigma_c^2 \end{bmatrix} \tag{3}$$

where $\mathbf{I}$ is the pixel color we actually measured. σ_c is the covariance of measurement noise for each channel.

In LVBA, we assume that all scene points lie on Lambertian surfaces, hence their radiance $\boldsymbol{\gamma}$ is the same in all directions. For two camera frames (i.e. the reference frame and target frames), when their poses are the ground truth and observing the same scene point, the observation radiance should be the same, which indicates:

$$\mathbf{0} = \boldsymbol{\gamma}_r - \boldsymbol{\gamma}_t = \tau_r^{-1}\mathbf{I}_r - \tau_t^{-1}\mathbf{I}_t + (\tau_t^{-1}\boldsymbol{\delta}_{ct} - \tau_r^{-1}\boldsymbol{\delta}_{cr}) \tag{4}$$

where $\boldsymbol{\gamma}_r$ and $\boldsymbol{\gamma}_t$ are the rays projected to the reference and target frames, respectively. τ_r and τ_t are the exposure times of the reference and target frames. $\mathbf{I}_r$ and $\mathbf{I}_t$ are the corresponding pixel colors when the rays are captured by the camera. $\boldsymbol{\delta}_{cr}$ and $\boldsymbol{\delta}_{ct}$ are the respective measurement noise. Thus we define the error function $\mathbf{e}$ as:

$$\mathbf{e} \coloneqq \tau_r^{-1}\mathbf{I}_r - \tau_t^{-1}\mathbf{I}_t \sim \mathcal{N}\left(0, \left(\frac{1}{\tau_t^2} + \frac{1}{\tau_r^2}\right)\boldsymbol{\Sigma}_c\right) \tag{5}$$

The cost item L can be expressed as:

$$L = \|\mathbf{e}\|^2_{(\tau_t^{-2}+\tau_r^{-2})\boldsymbol{\Sigma}_c} \tag{6}$$

The cost function will be the summation of all generated cost items. It should be noted that if we use (6) to construct the cost function and optimize it, the degrees of freedom for the exposure time is not sufficiently constrained. Within a sequence, there is always a general solution: when all exposure time approaches infinity, all the cost items L will tend to zero. To address this problem, we define a "Relative Exposure Time (RET)" and optimize it instead of the exposure time. For the i-th camera frame with the exposure time τ_i, its relative exposure time is defined as:

$$\epsilon_i = \frac{\tau_i}{\tau_1} \quad (i = r, t) \tag{7}$$

where τ_1 is the exposure time of the first camera frame, and there is always $\epsilon_1 = 1$. Then, the error $\mathbf{e}$ could be written as

$$\mathbf{e} = \tau_r^{-1}\mathbf{I}_r - \tau_t^{-1}\mathbf{I}_t = \tau_1^{-1}\left(\epsilon_r^{-1}\mathbf{I}_r - \epsilon_t^{-1}\mathbf{I}_t\right) \tag{8}$$

Since the constant scale factor τ_1^{-1} only scaled the whole cost function and does not affect the optimal camera state during optimization. We simply ignore it, leading to the error $\mathbf{e}_\epsilon$:

$$\begin{aligned} \mathbf{e}_\epsilon &\coloneqq \tau_1\mathbf{e} = \epsilon_r^{-1}\mathbf{I}_r - \epsilon_t^{-1}\mathbf{I}_t \\ &\sim \mathcal{N}\left(0, \tau_1^2\left(\frac{1}{\epsilon_t^2} + \frac{1}{\epsilon_r^2}\right)\boldsymbol{\Sigma}_c\right) \end{aligned} \tag{9}$$

Also, since the value of τ_1 and σ_c in our covariance only scaled the whole cost function, which does not affect the optimal camera states during optimization, we simply set $\sigma_c\tau_1 = 1$. And the final covariance matrix $\boldsymbol{\Sigma}$ will be

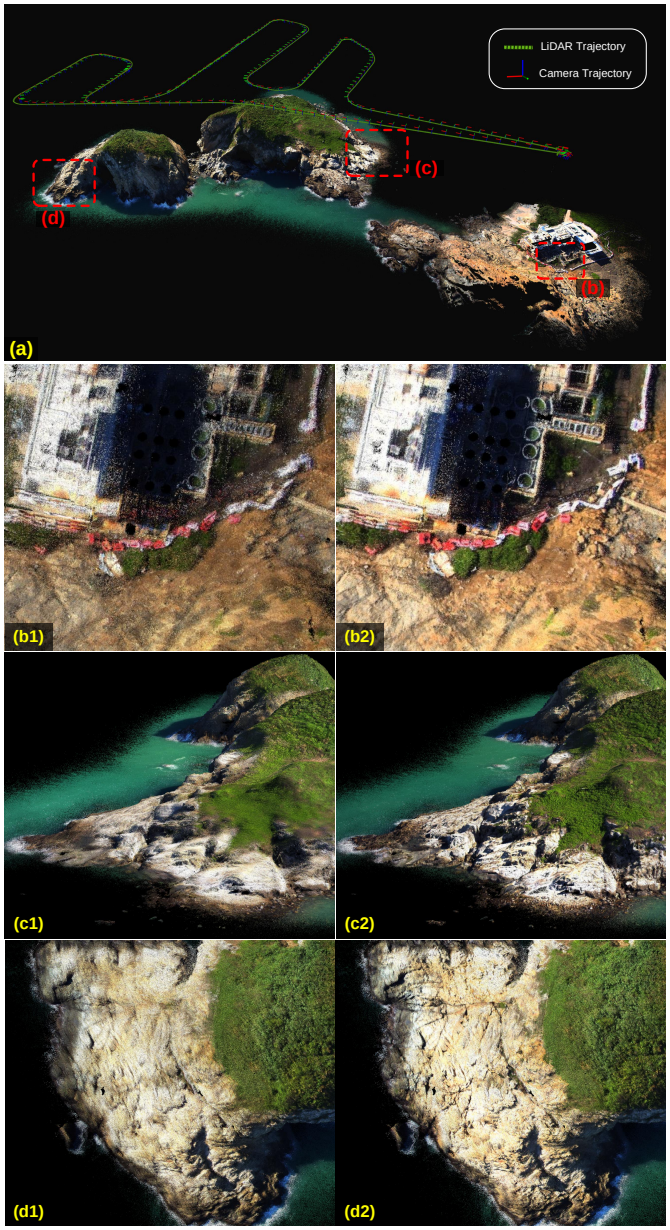
$$\boldsymbol{\Sigma} = \left(\frac{1}{\epsilon_t^2} + \frac{1}{\epsilon_r^2}\right)\mathbf{E} \tag{10}$$

where $\mathbf{E}$ is the 3×3 identity matrix. Combining (9) and (10), the final cost item is $\|\mathbf{e}_\epsilon\|^2_{\boldsymbol{\Sigma}}$ as illustrated in (8).

It is worthy mention that implementing the covariance (9) in the cost function $\|\epsilon_t^{-1}\mathbf{I}_t - \epsilon_r^{-1}\mathbf{I}_r\|^2$ is necessary. Otherwise, an apparently wrong optimal solution to the cost function would exist, which is $\epsilon_i = \infty, \forall i$. Incorporating the covariance in the cost function would rectify this issue by excluding the wrong optimum.

Appendix B
Mapping result for Asynchronous LiDAR and Camera Sequences

The results of our UAV mapping experiment. Panel (a) offers an overview of the outcome following our optimization process. Panels (b1), (c1), and (d1) show detailed views of the initial values before optimization. In contrast, panels (b2), (c2), and (d2) display the corresponding areas after our optimization.



REFERENCES

[1] J. Engel, V. Koltun, and D. Cremers, "Direct sparse odometry," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 3, pp. 611–625, 2017.