PRISM-VO: Scale-Aware Visual Odometry Using Photometric Plenoptic Bundle Adjustment Supplementary Material

Aymeric Fleith[1,2], Julian Zirbel[1,2], Daniel Cremers[1], and Niclas Zeller[2]

[1] Technical University of Munich, Munich, Germany
{aymeric.fleith, julian.zirbel, cremers}@tum.de
[2] Karlsruhe University of Applied Sciences, Karlsruhe, Germany
niclas.zeller@h-ka.de

A Introduction

This supplementary document provides additional details and derivations that complement the main paper. In particular, it presents the mathematical formulation of the plenoptic camera model, the projection functions used in our method, and the complete definition of the residuals and the Jacobian used in the proposed photometric plenoptic bundle adjustment.

Section B introduces the plenoptic camera model. Section C derives the projection functions, including the direct and inverse mappings between camera and virtual image coordinates, as well as projections between frames. Section D defines the photometric and inverse virtual depth residuals used in our optimization. Section E presents the analytical Jacobian with respect to affine brightness parameters, pose parameters, and inverse depth. Then, Section F describes the generation of RGB images compatible with monocular camera pipelines used to evaluate our method against monocular pipelines. Finally, Section G and Section H present additional results, comparing our method first to an RGB-D simultaneous localization and mapping (SLAM) algorithm and then to a monocular method using the original field of view of the images.

B Plenoptic Camera Model

The proposed pipeline is based on the plenoptic camera model introduced in [3], which combines a thin lens model for the main lens with a pinhole model for each micro-lens. In the model, we define the points in the following coordinate systems:

- $\mathbf{x}_C = [x_C, y_C, z_C]^T$ (or homogeneous coordinates $\overline{\mathbf{x}}_C = [x_C, y_C, z_C, 1]^T$): 3D point in the camera metric coordinate frame.
- $\mathbf{x}_V = [x_V, y_V, z_V]^T$ (or homogeneous coordinates $\overline{\mathbf{x}}_V = [x_V, y_V, z_V, 1]^T$): point in the metric virtual image space.
- $\mathbf{x}'_V = [x'_V, y'_V, z'_V]^T$ (or homogeneous coordinates $\overline{\mathbf{x}}'_V = [x'_V, y'_V, z'_V, 1]^T$): point in the dimensionless virtual image space.

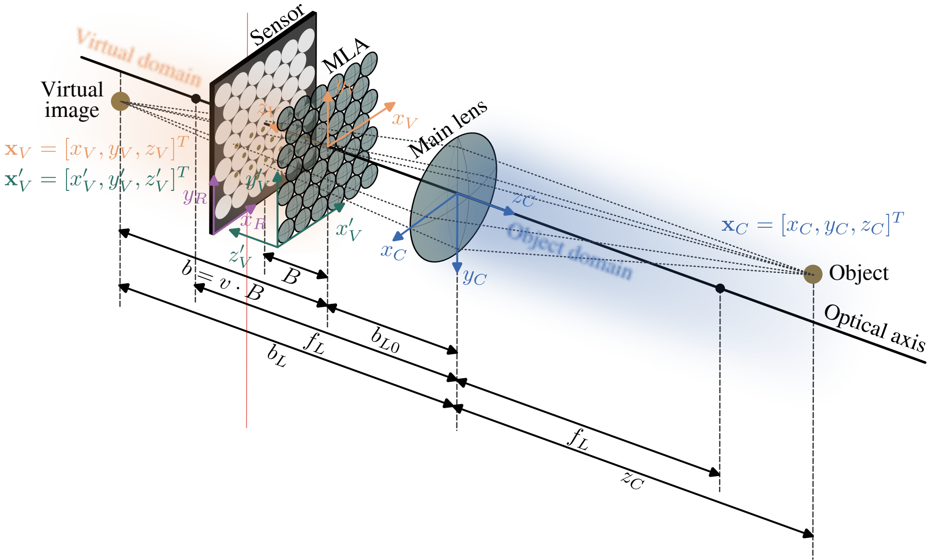

Fig. 1: Representation of the model of the focused plenoptic camera in Galilean mode inspired by the model in [3].

For simplicity, we do not distinguish between $\mathbf{x}_V = [x_V, y_V, z_V]^T$ and $\mathbf{x}'_V = [x'_V, y'_V, z'_V]^T$ in the main paper. The plenoptic camera model relies on the intrinsic parameters depicted in Fig. 1. The intrinsic camera parameters of the plenoptic model used are as follows:

- f_L: focal length of the main lens, defining the thin-lens geometry.
- $\mathbf{c}_L = [c_x, c_y]^T$: principal point offsets of the main lens.
- B: the distance between the micro-lens array (MLA) and the image sensor.
- b_{L0}: the distance between the main lens and the MLA.

Using the camera parameters, the projection from $\mathbf{x}_C$ to $\mathbf{x}_V$ is defined as

$$\lambda \cdot \overline{\mathbf{x}}_V = \mathbf{K} \cdot \overline{\mathbf{x}}_c \qquad \Leftrightarrow \qquad \lambda \cdot \begin{bmatrix} x_V \\ y_V \\ z_V \\ 1 \end{bmatrix} = \begin{bmatrix} b_L & 0 & 0 & 0 \\ 0 & b_L & 0 & 0 \\ 0 & 0 & b & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} x_C \\ y_C \\ z_C \\ 1 \end{bmatrix}. \tag{1}$$

The parameters b_L and b depend on the object point distance z_C. Using the thin lens equation, the parameter b_L is defined as

$$b_L = \left(\frac{1}{f_L} - \frac{1}{z_C} \right)^{-1} = b + b_{L0}. \tag{2}$$

The point $\mathbf{x}_V$ is defined in metric coordinates. It can be converted to a dimensionless point $\mathbf{x}'_V$ with Eq. (3) where the coordinates x'_V and y'_V are in pixels and the coordinate z'_V is defined by the virtual depth v introduced in [5] and defined in Eq. (4). The projection from $\mathbf{x}_V$ to $\mathbf{x}'_V$ uses the pixel sizes s_x and s_y to provide a metric reference for the parameters.

$$x'_V = x_V \cdot s_x^{-1} + c_x, \qquad y'_V = y_V \cdot s_y^{-1} + c_y \tag{3}$$

$$v = \frac{b}{B} = \frac{b_L - b_{L0}}{B} \quad \Rightarrow \quad b_L = v \cdot B + b_{L0} \tag{4}$$

We rewrite this projection as a matrix $\mathbf{K}_S$ in Eq. (5).

$$\overline{\mathbf{x}}'_V = \mathbf{K}_S \cdot \overline{\mathbf{x}}_V \qquad \Leftrightarrow \qquad \begin{bmatrix} x'_V \\ y'_V \\ z'_V \\ 1 \end{bmatrix} = \begin{bmatrix} s_x^{-1} & 0 & 0 & c_x \\ 0 & s_y^{-1} & 0 & c_y \\ 0 & 0 & B^{-1} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_V \\ y_V \\ z_V \\ 1 \end{bmatrix} \tag{5}$$

A virtual image point $\mathbf{x}'_V$ is then projected to the raw sensor image $\mathbf{x}_R = [x_R, y_R]^T$ through the corresponding micro-lens center $\mathbf{c}_{ML} = [c_{MLx}, c_{MLy}]^T$:

$$x_R = (x'_V - c_{MLx})\, v^{-1} + c_{MLx}, \tag{6}$$

$$y_R = (y'_V - c_{MLy})\, v^{-1} + c_{MLy}. \tag{7}$$

Micro-lens centers $\mathbf{c}_{ML}$ are derived from micro-image centers $\mathbf{c}_I = [c_{Ix}, c_{Iy}]^T$ via

$$\mathbf{c}_{ML} = \mathbf{c}_I \cdot \frac{b_{L0}}{b_{L0} + B}. \tag{8}$$

Lens distortion is applied directly to the raw image coordinates and micro-lens centers. The model incorporates radially symmetric and tangential distortion following [1], but can be replaced by any alternative distortion model.

C Plenoptic Camera Projection

The PRISM-VO pipeline uses a direct approach. Points present in keyframes are projected into the new frame to define a photometric residual based on the intensities of the points in the host frame and in the target frame. This section explains how to derive the projection from one frame to another based on the plenoptic camera model.

Based on Eq. (1), writing component-wise, we obtain Eq. (9).

$$\begin{cases} \lambda \cdot x_V = b_L \cdot x_C & \Leftrightarrow \quad x_V = \dfrac{x_C}{z_C} \cdot b_L \\ \lambda \cdot y_V = b_L \cdot y_C & \Leftrightarrow \quad y_V = \dfrac{y_C}{z_C} \cdot b_L \\ \lambda \cdot z_V = b \cdot z_C & \Leftrightarrow \quad z_V = b \\ \lambda = z_C & \end{cases} \tag{9}$$

C.1 Direct Mapping from $\mathbf{x}_C$ to $\mathbf{x}'_V$

The direct transformation from a point $\mathbf{x}_C$ in the camera coordinate frame to a point $\mathbf{x}'_V$ in the dimensionless virtual image space can be expressed directly by injecting Eq. (9) in Eq. (3). The forward plenoptic projection is fully defined by Eq. (10).

$$\begin{cases} x'_V = \dfrac{x_C}{z_C} \cdot b_L \cdot s_x^{-1} + c_x \\ y'_V = \dfrac{y_C}{z_C} \cdot b_L \cdot s_y^{-1} + c_y \\ z'_V = \dfrac{b_L - b_{L0}}{B} \end{cases} \tag{10}$$

C.2 Inverse Mapping from $\mathbf{x}'_V$ to $\mathbf{x}_C$

The inverse mapping from a virtual image point $\mathbf{x}'_V$ to a 3D point $\mathbf{x}_C$ in camera coordinates maps a point from the virtual image space of the camera to the world coordinate system. As an intermediate step, by inverting Eq. (3) we get Eq. (11).

$$\begin{cases} x_V = (x'_V - c_x)\, s_x \\ y_V = (y'_V - c_y)\, s_y \\ z_V = B \cdot v = B \cdot z'_V \end{cases} \tag{11}$$

By rearranging Eq. (2), we obtain the expression for z_C as a function of the camera parameters f_L, b, and b_{L0}. This new equation can be substituted into the expression for z_C in Eq. (9). Furthermore, by substituting Eq. (11), we finally obtain Eq. (12).

$$\begin{cases} x_C = \dfrac{\lambda \cdot x_V}{b_L} \\ y_C = \dfrac{\lambda \cdot y_V}{b_L} \\ z_C = \left(\dfrac{1}{f_L} - \dfrac{1}{b + b_{L0}} \right)^{-1} \end{cases} \Leftrightarrow \begin{cases} x_C = \dfrac{\lambda}{b_L} (x'_V - c_x)\, s_x \\ y_C = \dfrac{\lambda}{b_L} (y'_V - c_y)\, s_y \\ z_C = \left(\dfrac{1}{f_L} - \dfrac{1}{b_L} \right)^{-1} \end{cases} \tag{12}$$

C.3 Projection from One Frame to Another

Given the analytical expressions for (i) projecting a virtual point into a metric point in the camera coordinate system (Sec. C.2), and (ii) back-projecting a metric point from the camera coordinate system into the virtual domain (Sec. C.1), we can compose these mappings to formally define the transformation of a point $\mathbf{x}'_{V,i}$ from one reference host frame I_i to a point $\mathbf{x}'_{V,j}$ from a target frame I_j. The projection is illustrated in Fig. 2.

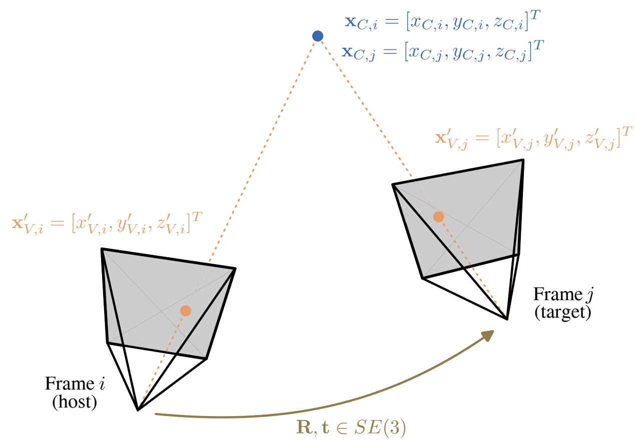

Fig. 2: Projection of a point $\mathbf{x}'_{V,i}$ from the image of frame I_i (host frame) to the corresponding point $\mathbf{x}'_{V,j}$ in the image of frame I_j (target frame) via the 3D point $\mathbf{x}'_C$.

First, we use Eq. (1) to express the virtual point $\overline{\mathbf{x}}_{V,j}$ from the frame I_j as a function of the 3D point $\overline{\mathbf{x}}_{C,j}$ in the camera coordinate system. We denote by $\mathbf{K}_j$ the corresponding conversion camera matrix from $\overline{\mathbf{x}}_{C,j}$ to $\overline{\mathbf{x}}_{V,j}$ for the frame I_j. The relation is given in Eq. (13).

$$\lambda_j \cdot \overline{\mathbf{x}}_{V,j} = \mathbf{K}_j \cdot \overline{\mathbf{x}}_{C,j} \qquad \Leftrightarrow \qquad \overline{\mathbf{x}}_{V,j} = \frac{1}{\lambda_j} \cdot \mathbf{K}_j \cdot \overline{\mathbf{x}}_{C,j}. \tag{13}$$

Let $\mathbf{K}_{S,j}$ denote the conversion matrix between the image point $\overline{\mathbf{x}}_{V,j}$ in the virtual image space and the image point $\overline{\mathbf{x}}'_{V,j}$ in the dimensionless virtual image space for the frame I_j. So, from Eq. (5), we have

$$\overline{\mathbf{x}}'_{V,j} = \mathbf{K}_{S,j} \cdot \overline{\mathbf{x}}_{V,j}. \tag{14}$$

We set $\mathbf{R}, \mathbf{t} \in SE(3)$ as the transformation between the poses of the host frame I_i and the target frame I_j. We can establish a relationship between the 3D point $\mathbf{x}_{C,i}$ observed in frame I_i and the point $\mathbf{x}_{C,j}$ observed in frame I_j with Eq. (15).

$$\overline{\mathbf{x}}_{C,j} = \mathbf{R} \cdot \overline{\mathbf{x}}_{C,i} + \mathbf{t} \tag{15}$$

An equality can then be established between Eq. (13) and Eq. (14). By also substituting Eq. (15) we get

$$\frac{1}{\lambda_j} \cdot \mathbf{K}_j \cdot \overline{\mathbf{x}}_{C,j} = \mathbf{K}_{S,j}^{-1} \cdot \overline{\mathbf{x}}'_{V,j} \qquad \Leftrightarrow \qquad \lambda_j \cdot \overline{\mathbf{x}}'_{V,j} = \mathbf{K}_{S,j} \cdot \mathbf{K}_j \cdot (\mathbf{R} \cdot \overline{\mathbf{x}}_{C,i} + \mathbf{t}) \tag{16}$$

with

$$\bar{\mathbf{x}}_{C,i} = \begin{bmatrix} x_{C,i} \\ y_{C,i} \\ z_{C,i} \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{z_{C,i}}{b_{L,i}} \left(x'_{V,i} - c_x\right) s_x \\ \frac{z_{C,i}}{b_{L,i}} \left(y'_{V,i} - c_y\right) s_y \\ \left(\frac{1}{f_L} - \frac{1}{b_{L,i}}\right)^{-1} \\ 1 \end{bmatrix} . \tag{17}$$

Finally, the projection of a point $\mathbf{x}'_{V,i} = [x'_{V,i}, y'_{V,i}, z'_{V,i}]^T$ in frame I_i to a point $\mathbf{x}'_{V,j} = [x'_{V,j}, y'_{V,j}, z'_{V,j}]^T$ in frame I_j is defined as

$$\mathbf{x}'_{V,j} = \Pi_{\text{pl}}\left(\mathbf{x}'_{V,i}, z_{C,i}; \mathbf{R}, \mathbf{t}, \boldsymbol{\theta}\right) , \tag{18}$$

where $\boldsymbol{\theta} = \{f_L, \mathbf{c}_L, B, b_{L0}\}$ is the set of camera parameters and $z_{C,i}$ the depth of the point along the optical axis. In a developed form, the projection $\Pi_{\text{pl}}(\cdot)$ is given in Eq. (19). For ease of notation, the projection will be referred to as $\Pi_{\text{pl}}(\mathbf{x}'_{V,i})$ subsequently.

$$\lambda_j \begin{bmatrix} x'_{V,j} \\ y'_{V,j} \\ z'_{V,j} \\ 1 \end{bmatrix} = \mathbf{K}_{S,j}\mathbf{K}_j \left(\mathbf{R} \cdot \begin{bmatrix} \frac{z_{C,i}}{b_{L,i}} (x'_{V,i} - c_x) s_x \\ \frac{z_{C,i}}{b_{L,i}} (y'_{V,i} - c_y) s_y \\ \left(\frac{1}{f_L} - \frac{1}{b_{L,i}}\right)^{-1} \\ 1 \end{bmatrix} + \mathbf{t} \right) \tag{19}$$

D Residuals Definition

The PRISM-VO pipeline defines both photometric residuals and inverse virtual depth residuals. The photometric residual compares intensity values in the host frame and the target frame. The inverse virtual depth residual compares the estimated depth data with the measured depth data.

D.1 Photometric Residual

The photometric residual is defined as the difference between two image intensities. On the one hand, we take the intensity of a point $\mathbf{x}_{V,i}$ from a host frame I_i. This point is projected into a target frame I_j by the projection denoted $\Pi_{\text{pl}}(\cdot)$ and defined in Eq. (19). This allows, on the other hand, to consider the intensity of the position reached after projection. The difference between both intensities is minimized. The photometric residual r^{photo} for a point $\mathbf{x}_{V,i}$ in a reference frame I_i observed in a target frame I_j is defined as

$$r^{\text{photo}} = \left(I_j\left[\Pi_{\text{pl}}(\mathbf{x}'_{V,i})\right] - b_j\right) - \frac{e^{a_j}}{e^{a_i}}\left(I_i[\mathbf{x}'_{V,i}] - b_i\right) . \tag{20}$$

The residual is simplified and reformulated to reveal the affine brightness parameters a and b. The reformulation is given in Eq. (21).

$$
\begin{aligned}
r^{\text{photo}} &= \left(I_j\left[\mathbf{x}'_{V,j}\right] - b_j\right) - \frac{e^{a_j}}{e^{a_i}}\left(I_i[\mathbf{x}'_{V,i}] - b_i\right) \\
&= I_j\left[\mathbf{x}'_{V,j}\right] - b_j - e^{a_j - a_i}\left(I_i[\mathbf{x}'_{V,i}]] - b_i\right) \\
&= I_j\left[\mathbf{x}'_{V,j}\right] - \underbrace{e^{a_j - a_i}}_{\text{a}} I_i[\mathbf{x}'_{V,i}] + \underbrace{e^{a_j - a_i}}_{\text{a}} b_i - b_j \\
&= I_j\left[\mathbf{x}'_{V,j}\right] - (aI_i[\mathbf{x}'_{V,i}] + b_j - \underbrace{ab_i}_{\text{b}}) \\
&= I_j\left[\mathbf{x}'_{V,j}\right] - \left(aI_i[\mathbf{x}'_{V,i}] + b_j - b\right)
\end{aligned} \tag{21}
$$

D.2 Inverse Virtual Depth Residual

The depth residual r^{depth} is defined in a single frame in the inverse virtual depth domain. It compares, on the one hand, the inverse virtual depth ρ_v^{mes} calculated by a single camera view. On the other hand, we take the inverse virtual depth ρ_v estimated in the pipeline. The residual is set as in Eq. (22).

$$
r^{\text{depth}} = \rho_v - \rho_v^{\text{mes}} = \frac{1}{v} - \frac{1}{v^{\text{mes}}} \tag{22}
$$

E Jacobian Definition

As explained in the main paper, the nonlinear formulation is optimized using a Levenberg-Marquardt algorithm. The optimization jointly estimates the camera poses $\boldsymbol{\xi}_i \in SE(3)$, the affine brightness parameters a_i and b_i, and the inverse depth values $\rho_{C,i}$. Considering the photometric residuals (Sec. D.1) and inverse virtual depth residuals (Sec. D.2), the Jacobian is composed for each frame of the partial derivatives given in Eq. (23).

$$
\begin{aligned}
&\frac{\partial r^{\text{photo}}}{\partial a_i}; \quad \frac{\partial r^{\text{photo}}}{\partial b_i}; \quad \frac{\partial r^{\text{photo}}}{\partial \boldsymbol{\xi}_i}; \quad \frac{\partial r^{\text{photo}}}{\partial \rho_{C,i}}; \\
&\frac{\partial r^{\text{depth}}}{\partial a_i}; \quad \frac{\partial r^{\text{depth}}}{\partial b_i}; \quad \frac{\partial r^{\text{depth}}}{\partial ; \boldsymbol{\xi}_i}; \quad \frac{\partial r^{\text{depth}}}{\partial \rho_{C,i}}
\end{aligned} \tag{23}
$$

This section explains how the different derivatives of the matrix are obtained.

E.1 Affine Parameter Derivatives

Affine Parameter Derivatives of the Photometric Residual. We calculate the derivative of the residual in Eq. (21). Its derivative with respect to the parameter a is:

$$\begin{aligned}\frac{\partial r^{\text{photo}}}{\partial a} &= \frac{\partial}{\partial a}\left(I_j\left[\mathbf{x}'_{V,j}\right] - \left(aI_i[\mathbf{x}'_{V,i}] + b_j - b\right)\right) \\ &= \frac{\partial}{\partial a}\left(I_j\left[\mathbf{x}'_{V,j}\right] - \left(aI_i[\mathbf{x}'_{V,i}] + b_j - ab_i\right)\right) \\ &= b_i - I_i[\mathbf{x}'_{V,i}].\end{aligned} \tag{24}$$

We proceed similarly for the parameter b:

$$\frac{\partial r^{\text{photo}}}{\partial b} = \frac{\partial}{\partial a}\left(I_j\left[\mathbf{x}'_{V,j}\right] - \left(aI_i[\mathbf{x}'_{V,i}] + b_j - b\right)\right) = 1. \tag{25}$$

Affine Parameter Derivatives of the Depth Residual. The parameters a and b are specific to the photometric residual. Thus, we can directly establish the relationships in Eq. (26).

$$\frac{\partial r^{\text{depth}}}{\partial a} = \frac{\partial r^{\text{depth}}}{\partial b} = 0 \tag{26}$$

E.2 Pose Parameter Derivatives

The poses of the different camera views are denoted by $\boldsymbol{\xi}_i$ for the pose of frame I_i. We represent the pose as the 6-dimensional vector described by Eq. (27). $\mathbf{t} = [t_1, t_2, t_3]^T \in \mathbb{R}^3$ denotes the linear component and $\boldsymbol{\omega} = [\omega_1, \omega_2, \omega_3]^T \in \mathbb{R}^3$ denotes the angular component of the pose. This vector representation corresponds to an element of the Lie algebra $\mathfrak{se}(3)$ associated with the Special Euclidean group $SE(3)$.

$$\boldsymbol{\xi}_i = \begin{bmatrix}\mathbf{t} \\ \boldsymbol{\omega}\end{bmatrix} = \left[t_1, t_2, t_3, \omega_1, \omega_2, \omega_3\right]^T \in \mathbb{R}^6 \tag{27}$$

Pose Parameter Derivatives of the Photometric Residual. The derivative of the photometric residual as a function of camera poses $\boldsymbol{\xi}_i$ can be decomposed using the chain rule. Thus, we can write:

$$\frac{\partial r^{\text{photo}}}{\partial \boldsymbol{\xi}_i} = \frac{\partial r^{\text{photo}}}{\partial \mathbf{x}'_{V,j}} \cdot \frac{\partial \mathbf{x}'_{V,j}}{\partial \mathbf{x}_{V,j}} \cdot \frac{\partial \mathbf{x}_{V,j}}{\partial \mathbf{x}_{C,j}} \cdot \frac{\partial \mathbf{x}_{C,j}}{\partial \boldsymbol{\xi}_i}. \tag{28}$$

Based on the decomposition in Eq. (28), we can first calculate the different parts of the chain independently.

- Calculation of $\frac{\partial r^{\text{photo}}}{\partial \mathbf{x}'_{V,j}}$

Using the definition in Eq. (21), we can establish the derivative of the photometric residual as a function of the virtual point $\mathbf{x}'_{V,j}$:

$$\begin{aligned}\frac{\partial r^{\text{photo}}}{\partial \mathbf{x}'_{V,j}} &= \frac{\partial}{\partial \mathbf{x}'_{V,j}} \left(I_j\left[\mathbf{x}'_{V,j}\right] - \left(aI_i[\mathbf{x}'_{V,i}] + b_j - b\right)\right) \\ &= \frac{\partial}{\partial \mathbf{x}'_{V,j}} \left(I_j\left[\mathbf{x}'_{V,j}\right]\right) \\ &= \begin{bmatrix} d_x \\ d_y \end{bmatrix}^T . \end{aligned} \tag{29}$$

Note that the equation Eq. (29) actually represents the gradient of the image at point $\mathbf{x}_{V,j}$.

– Calculation of $\frac{\partial \mathbf{x}'_{V,j}}{\partial \mathbf{x}_{V,j}}$

The expression of $\mathbf{x}'_{V,j}$ as a function of $\mathbf{x}_{V,j}$ is given by reformulating Eq. (11). We can then define the full derivative matrix by Eq. (30).

$$\frac{\partial \mathbf{x}'_{V,j}}{\partial x_{V,j}} = \begin{bmatrix} \frac{\partial x'_{V,j}}{\partial x_{V,j}} & \frac{\partial x'_{V,j}}{\partial y_{V,j}} & \frac{\partial x'_{V,j}}{\partial z_{V,j}} \\ \frac{\partial y'_{V,j}}{\partial x_{V,j}} & \frac{\partial y'_{V,j}}{\partial y_{V,j}} & \frac{\partial y'_{V,j}}{\partial z_{V,j}} \\ \frac{\partial z'_{V,j}}{\partial x_{V,j}} & \frac{\partial z'_{V,j}}{\partial y_{V,j}} & \frac{\partial z'_{V,j}}{\partial z_{V,j}} \end{bmatrix} \tag{30}$$

We compute the individual components of the matrix in Eq. (30).

$$\frac{\partial x'_{V,j}}{\partial x_{V,j}} = \frac{\partial}{\partial x_{V,j}} \left(x_{V,j} \cdot s_x^{-1} + c_x\right) = s_x^{-1} \tag{31}$$

$$\frac{\partial y'_{V,j}}{\partial y_{V,j}} = \frac{\partial}{\partial y_{V,j}} \left(y_{V,j} \cdot s_y^{-1} + c_y\right) = s_y^{-1} \tag{32}$$

$$\frac{\partial z'_{V,j}}{\partial z_{V,j}} = \frac{\partial}{\partial z_{V,j}} \left(z_{V,j} \cdot B^{-1}\right) = B^{-1} \tag{33}$$

$$\frac{\partial x'_{V,j}}{\partial y_{V,j}} = \frac{\partial x'_{V,j}}{\partial z_{V,j}} = \frac{\partial y'_{V,j}}{\partial x_{V,j}} = \frac{\partial y'_{V,j}}{\partial z_{V,j}} = \frac{\partial z'_{V,j}}{\partial x_{V,j}} = \frac{\partial z'_{V,j}}{\partial y_{V,j}} = 0 \tag{34}$$

So, finally we get:

$$\frac{\partial \mathbf{x}'_{V,j}}{\partial \mathbf{x}_{V,j}} = \begin{bmatrix} s_x^{-1} & 0 & 0 \\ 0 & s_y^{-1} & 0 \\ 0 & 0 & B^{-1} \end{bmatrix} . \tag{35}$$

– Calculation of $\frac{\partial \mathbf{x}_{V,j}}{\partial \mathbf{x}_{C,j}}$

The expression of $\mathbf{x}_{V,j}$ as a function of $\mathbf{x}_{C,j}$ is defined by Eq. (13) by associating the components of Eq. (1). To simplify notation, we define the inverse metric depth $\rho_{C,j}$ of the image point in frame I_j. Thus, we have the relation in Eq. (36).

$$\rho_{C,j} = \frac{1}{z_{C,j}} = \frac{1}{\lambda_j} \tag{36}$$

Using the relationship for b_2 from Eq. (4), we can write:

$$\mathbf{x}_{V,j} = \begin{bmatrix} x_{V,j} \\ y_{V,j} \\ z_{V,j} \\ 1 \end{bmatrix} = \rho_{C,j} \cdot \begin{bmatrix} b_{L,j} \cdot x_{C,j} \\ b_{L,j} \cdot y_{C,j} \\ b_2 \cdot z_{C,j} \\ z_{C,j} \end{bmatrix} = \rho_{C,j} \cdot \begin{bmatrix} b_{L,j} \cdot x_{C,j} \\ b_{L,j} \cdot y_{C,j} \\ (b_{L,j} - b_{L0}) \cdot z_{C,j} \\ z_{C,j} \end{bmatrix}. \tag{37}$$

We can then simplify the expression of Eq. (2) for $b_{L,j}$ and inject it into Eq. (37) in order to describe $\mathbf{x}_{V,j}$ as a function of $\mathbf{x}_{C,j}$, as shown in Eq. (38).

$$\overline{X}_{V,j} = \begin{bmatrix} x_{V,j} \\ y_{V,j} \\ z_{V,j} \\ 1 \end{bmatrix} = \rho_{C,j} \cdot \begin{bmatrix} \frac{f_L \cdot z_{C,j}}{z_{C,j} - f_L} \cdot x_{C,j} \\ \frac{f_L \cdot z_{C,j}}{z_{C,j} - f_L} \cdot y_{C,j} \\ \left(\frac{f_L \cdot z_{C,j}}{z_{C,j} - f_L} - b_{L0}\right) \cdot z_{C,j} \\ z_{C,j} \end{bmatrix} \tag{38}$$

Finally, the derivative matrix to be calculated is the one given in Eq. (39), considering the expressions previously demonstrated and set out in Eq. (38).

$$\frac{\partial \mathbf{x}_{V,j}}{\partial \mathbf{x}_{C,j}} = \begin{bmatrix} \frac{\partial x_{V,j}}{\partial x_{C,j}} & \frac{\partial x_{V,j}}{\partial y_{C,j}} & \frac{\partial x_{V,j}}{\partial z_{C,j}} \\ \frac{\partial y_{V,j}}{\partial x_{C,j}} & \frac{\partial y_{V,j}}{\partial y_{C,j}} & \frac{\partial y_{V,j}}{\partial z_{C,j}} \\ \frac{\partial z_{V,j}}{\partial x_{C,j}} & \frac{\partial z_{V,j}}{\partial y_{C,j}} & \frac{\partial z_{V,j}}{\partial z_{C,j}} \end{bmatrix} \tag{39}$$

From this, we calculate the different components of the matrix in Eq. (39).

$$\begin{aligned} \frac{\partial x_{V,j}}{\partial x_{C,j}} &= \frac{\partial}{\partial x_{C,j}} \left(\rho_{C,j} \cdot \frac{f_L \cdot z_{C,j}}{z_{C,j} - f_L} \cdot x_{C,j}\right) = \frac{\partial}{\partial x_{C,j}} \left(\frac{1}{z_{C,j}} \cdot \frac{f_L \cdot z_{C,j}}{z_{C,j} - f_L} \cdot x_{C,j}\right) \\ &= \frac{\partial}{\partial x_{C,j}} \left(\frac{f_L}{z_{C,j} - f_L} \cdot x_{C,j}\right) = \frac{f_L}{z_{C,j} - f_L} = \frac{f_L}{\frac{1}{\rho_{C,j}} - f_L} \end{aligned} \tag{40}$$

$$
\begin{aligned}
\frac{\partial x_{V,j}}{\partial z_{C,j}} &= \frac{\partial}{\partial z_{C,j}} \left(\rho_{C,j} \cdot \frac{f_L \cdot z_{C,j}}{z_{C,j} - f_L} \cdot x_{C,j} \right) = \frac{\partial}{\partial z_{C,j}} \left(\frac{1}{z_{C,j}} \cdot \frac{f_L \cdot z_{C,j}}{z_{C,j} - f_L} \cdot x_{C,j} \right) \\
&= \frac{\partial}{\partial z_{C,j}} \left(\frac{f_L}{z_{C,j} - f_L} \cdot x_{C,j} \right) = x_{C,j} \cdot \frac{\partial}{\partial z_{C,j}} \left(\frac{f_L}{z_{C,j} - f_L} \right) \\
&= -x_{C,j} \cdot \frac{f_L}{\left(z_{C,j} - f_L\right)^2} = -x_{C,j} \cdot \frac{f_L}{\left(\frac{1}{\rho_{C,j}} - f_L\right)^2}
\end{aligned}
\tag{41}
$$

$$
\begin{aligned}
\frac{\partial y_{V,j}}{\partial y_{C,j}} &= \frac{\partial}{\partial y_{C,j}} \left(\rho_{C,j} \cdot \frac{f_L \cdot z_{C,j}}{z_{C,j} - f_L} \cdot y_{C,j} \right) = \frac{\partial}{\partial y_{C,j}} \left(\frac{1}{z_{C,j}} \cdot \frac{f_L \cdot z_{C,j}}{z_{C,j} - f_L} \cdot y_{C,j} \right) \\
&= \frac{\partial}{\partial y_{C,j}} \left(\frac{f_L}{z_{C,j} - f_L} \cdot y_{C,j} \right) = \frac{f_L}{z_{C,j} - f_L} = \frac{f_L}{\frac{1}{\rho_{C,j}} - f_L}
\end{aligned}
\tag{42}
$$

$$
\begin{aligned}
\frac{\partial y_{V,j}}{\partial z_{C,j}} &= \frac{\partial}{\partial z_{C,j}} \left(\rho_{C,j} \cdot \frac{f_L \cdot z_{C,j}}{z_{C,j} - f_L} \cdot y_{C,j} \right) = \frac{\partial}{\partial z_{C,j}} \left(\frac{1}{z_{C,j}} \cdot \frac{f_L \cdot z_{C,j}}{z_{C,j} - f_L} \cdot y_{C,j} \right) \\
&= \frac{\partial}{\partial z_{C,j}} \left(\frac{f_L}{z_{C,j} - f_L} \cdot y_{C,j} \right) = y_{C,j} \cdot \frac{\partial}{\partial z_{C,j}} \left(\frac{f_L}{z_{C,j} - f_L} \right) \\
&= -y_{C,j} \cdot \frac{f_L}{\left(z_{C,j} - f_L\right)^2} = -y_{C,j} \cdot \frac{f_L}{\left(\frac{1}{\rho_{C,j}} - f_L\right)^2}
\end{aligned}
\tag{43}
$$

$$
\begin{aligned}
\frac{\partial z_{V,j}}{\partial z_{C,j}} &= \frac{\partial}{\partial z_{C,j}} \left(\rho_{C,j} \cdot \left(\frac{f_L \cdot z_{C,j}}{z_{C,j} - f_L} - b_{L0} \right) \cdot z_{C,j} \right) \\
&= \frac{\partial}{\partial z_{C,j}} \left(\frac{1}{z_{C,j}} \cdot \left(\frac{f_L \cdot z_{C,j}}{z_{C,j} - f_L} - b_{L0} \right) \cdot z_{C,j} \right) \\
&= \frac{\partial}{\partial z_{C,j}} \left(\frac{f_L \cdot z_{C,j}}{z_{C,j} - f_L} - b_{L0} \right) = \frac{\partial}{\partial z_{C,j}} \left(\frac{f_L \cdot z_{C,j}}{z_{C,j} - f_L} \right) \\
&= \frac{f_L \cdot (z_{C,j} - f_L) - f_L \cdot z_{C,j}}{\left(z_{C,j} - f_L\right)^2} = -\frac{f_L^2}{\left(z_{C,j} - f_L\right)^2} = -\frac{f_L^2}{\left(\frac{1}{\rho_{C,j}} - f_L\right)^2}
\end{aligned}
\tag{44}
$$

$$
\frac{\partial x_{V,j}}{\partial y_{C,j}} = \frac{\partial y_{V,j}}{\partial x_{C,j}} = \frac{\partial z_{V,j}}{\partial x_{C,j}} = \frac{\partial z_{V,j}}{\partial y_{C,j}} = 0 \tag{45}
$$

As a result:

$$\frac{\partial \mathbf{x}_{V,j}}{\partial \mathbf{x}_{C,j}} = \begin{bmatrix} \frac{f_L}{\frac{1}{\rho_{C,j}} - f_L} & 0 & -x_{C,j} \cdot \frac{f_L}{\left(\frac{1}{\rho_{C,j}} - f_L\right)^2} \\ 0 & \frac{f_L}{\frac{1}{\rho_{C,j}} - f_L} & -y_{C,j} \cdot \frac{f_L}{\left(\frac{1}{\rho_{C,j}} - f_L\right)^2} \\ 0 & 0 & -\frac{f_L^2}{\left(\frac{1}{\rho_{C,j}} - f_L\right)^2} \end{bmatrix} \tag{46}$$

We can now calculate $\frac{\partial \mathbf{x}'_{V,j}}{\partial X_{C,j}}$ in Eq. (47) by multiplying Eq. (35) and Eq. (46).

$$\begin{aligned} \frac{\partial \mathbf{x}'_{V,j}}{\partial \mathbf{x}_{C,j}} &= \begin{bmatrix} s_x^{-1} & 0 & 0 \\ 0 & s_y^{-1} & 0 \\ 0 & 0 & B^{-1} \end{bmatrix} \cdot \begin{bmatrix} \frac{f_L}{\frac{1}{\rho_{C,j}} - f_L} & 0 & -x_{C,j} \cdot \frac{f_L}{\left(\frac{1}{\rho_{C,j}} - f_L\right)^2} \\ 0 & \frac{f_L}{\frac{1}{\rho_{C,j}} - f_L} & -y_{C,j} \cdot \frac{f_L}{\left(\frac{1}{\rho_{C,j}} - f_L\right)^2} \\ 0 & 0 & -\frac{f_L^2}{\left(\frac{1}{\rho_{C,j}} - f_L\right)^2} \end{bmatrix} \\ &= \begin{bmatrix} \frac{f_L}{s_x \cdot \left(\frac{1}{\rho_{C,j}} - f_L\right)} & 0 & -x_{C,j} \cdot \frac{f_L}{s_x \cdot \left(\frac{1}{\rho_{C,j}} - f_L\right)^2} \\ 0 & \frac{f_L}{s_y \cdot \left(\frac{1}{\rho_{C,j}} - f_L\right)} & -y_{C,j} \cdot \frac{f_L}{s_y \cdot \left(\frac{1}{\rho_{C,j}} - f_L\right)^2} \\ 0 & 0 & -\frac{f_L^2}{B \cdot \left(\frac{1}{\rho_{C,j}} - f_L\right)^2} \end{bmatrix} \end{aligned} \tag{47}$$

– Calculation of $\frac{\partial \mathbf{x}_{C,j}}{\partial \boldsymbol{\xi}_i}$

To exploit the Lie group structure in matrix form, we define the hat operator $\hat{\cdot} : \mathbb{R}^6 \rightarrow \mathfrak{se}(3)$, which maps the pose vector to its matrix representation. Here, $\hat{\boldsymbol{w}} \in \mathbb{R}^{3\times 3}$ is the skew-symmetric matrix associated with the angular vector $\boldsymbol{w}$. Thus, the full matrix representation of the pose becomes Eq. (48).

$$\hat{\boldsymbol{\xi}}_i = \begin{bmatrix} \hat{\boldsymbol{w}} & \boldsymbol{v} \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & -\omega_3 & \omega_2 & t_1 \\ \omega_3 & 0 & -\omega_1 & t_2 \\ -\omega_2 & \omega_1 & 0 & t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{48}$$

For a small motion increment, we define the corresponding transformation update using the exponential map of the Lie algebra element. Given a perturbation $\delta\boldsymbol{\xi}_i \in \mathfrak{se}(3)$, the incremental rigid-body transformation is

$$\Delta\mathbf{T} = e^{\delta\boldsymbol{\xi}_i} \in SE(3). \tag{49}$$

This formulation ensures that the update remains on the manifold $SE(3)$, making it suitable for iterative optimization and state estimation on Lie groups. We compute the Jacobian of $\mathbf{x}_{C,j}$ with respect to the pose parameters $\boldsymbol{\xi}_i$ using a first-order approximation of the exponential map on $SE(3)$. For an infinitesimal perturbation of $\boldsymbol{\xi}$, the exponential map admits the first-order approximation

$$\exp(\hat{\boldsymbol{\xi}}_i) \approx I + \hat{\boldsymbol{\xi}}_i. \tag{50}$$

Thus, the derivative is obtained via the limit definition (see Eq. (51)).

$$\begin{aligned}
\frac{\partial \mathbf{x}_{C,j}}{\partial \boldsymbol{\xi}_i} &= \lim_{\boldsymbol{\xi}_i \to 0} \frac{e^{\hat{\boldsymbol{\xi}}_i} \cdot \mathbf{x}_{C,j} - \mathbf{x}_{C,j}}{\boldsymbol{\xi}_i} = \lim_{\boldsymbol{\xi}_i \to 0} \frac{\left(I + \hat{\boldsymbol{\xi}}_i\right) \cdot \mathbf{x}_{C,j} - \mathbf{x}_{C,j}}{\boldsymbol{\xi}_i} \\
&= \lim_{\boldsymbol{\xi}_i \to 0} \frac{\hat{\boldsymbol{\xi}}_i \cdot \mathbf{x}_{C,j}}{\boldsymbol{\xi}_i} = \lim_{\boldsymbol{\xi}_i \to 0} \frac{\begin{bmatrix} \hat{w} & v \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} \mathbf{x}_{C,j} \\ 1 \end{bmatrix}}{\boldsymbol{\xi}_i} \\
&= \lim_{\boldsymbol{\xi}_i \to 0} \frac{1}{\boldsymbol{\xi}_i} \cdot \begin{bmatrix} \hat{w} & v \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} \mathbf{x}_{C,j} \\ 1 \end{bmatrix} \\
&= \lim_{\boldsymbol{\xi}_i \to 0} \frac{1}{\boldsymbol{\xi}_i} \cdot \begin{bmatrix} 0 & -\omega_3 & \omega_2 & t_1 \\ \omega_3 & 0 & -\omega_1 & t_2 \\ -\omega_2 & \omega_1 & 0 & t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_{C,j} \\ y_{C,j} \\ z_{C,j} \\ 1 \end{bmatrix} \\
&= \lim_{\boldsymbol{\xi}_i \to 0} \frac{1}{\boldsymbol{\xi}_i} \cdot \begin{bmatrix} -\omega_3 \cdot y_{C,j} + \omega_2 \cdot z_{C,j} + t_1 \\ \omega_3 \cdot x_{C,j} - \omega_1 \cdot z_{C,j} + t_2 \\ -\omega_2 \cdot x_{C,j} + \omega_1 \cdot y_{C,j} + t_3 \\ 1 \end{bmatrix}
\end{aligned} \tag{51}$$

Computing the matrix components yields:

$$\frac{\partial \mathbf{x}_{C,j}}{\partial \boldsymbol{\xi}_i} = \begin{bmatrix} 1 & 0 & 0 & 0 & z_2 & -y_2 \\ 0 & 1 & 0 & -z_2 & 0 & x_2 \\ 0 & 0 & 1 & y_2 & -x_2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{52}$$

– Calculation of $\frac{\partial \mathbf{x}'_{V,j}}{\partial \boldsymbol{\xi}_i}$

Using the chain rule and the already calculated matrices, the Jacobian of the point $\mathbf{x}'_{V,j}$ with respect to the pose parameters $\boldsymbol{\xi}_i$ is given by Eq. (54).

$$\frac{\partial \mathbf{x}'_{V,j}}{\partial \boldsymbol{\xi}_i} = \frac{\partial \mathbf{x}'_{V,j}}{\partial \mathbf{x}_{C,j}} \cdot \frac{\partial \mathbf{x}_{C,j}}{\partial \boldsymbol{\xi}_i} \tag{53}$$

$$= \begin{bmatrix} \frac{f_L}{s_x \cdot \left(\frac{1}{\rho_{C,j}} - f_L\right)} & 0 & -x_{C,j} \cdot \frac{f_L}{s_x \cdot \left(\frac{1}{\rho_{C,j}} - f_L\right)^2} \\ 0 & \frac{f_L}{s_y \cdot \left(\frac{1}{\rho_{C,j}} - f_L\right)} & -y_{C,j} \cdot \frac{f_L}{s_y \cdot \left(\frac{1}{\rho_{C,j}} - f_L\right)^2} \\ 0 & 0 & -\frac{f_L^2}{B \cdot \left(\frac{1}{\rho_{C,j}} - f_L\right)^2} \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 & 0 & z_2 & -y_2 \\ 0 & 1 & 0 & -z_2 & 0 & x_2 \\ 0 & 0 & 1 & y_2 & -x_2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{54}$$

We introduce the following substitutions:

$$z_{C,j} = \frac{1}{\rho_{C,j}}, \tag{55}$$

$$u' = \frac{x_{C,j}}{z_{C,j}} \qquad \Leftrightarrow \qquad x_{C,j} = u' \cdot z_{C,j} = \frac{u'}{\rho_{C,j}}, \tag{56}$$

$$v' = \frac{y_{C,j}}{z_{C,j}} \qquad \Leftrightarrow \qquad y_{C,j} = v' \cdot z_{C,j} = \frac{v'}{\rho_{C,j}}. \tag{57}$$

Then, we calculate the different components.

$$\left.\frac{\partial \mathbf{x}'_{V,j}}{\partial \boldsymbol{\xi}_i}\right|_{(2,1)} = \left.\frac{\partial \mathbf{x}'_{V,j}}{\partial \boldsymbol{\xi}_i}\right|_{(3,1)} = \left.\frac{\partial \mathbf{x}'_{V,j}}{\partial \boldsymbol{\xi}_i}\right|_{(1,2)} = \left.\frac{\partial \mathbf{x}'_{V,j}}{\partial \boldsymbol{\xi}_i}\right|_{(3,2)} = \left.\frac{\partial \mathbf{x}'_{V,j}}{\partial \boldsymbol{\xi}_i}\right|_{(3,6)} = 0 \tag{58}$$

$$\left.\frac{\partial \mathbf{x}'_{V,j}}{\partial \boldsymbol{\xi}_i}\right|_{(1,1)} = \frac{f_L}{s_x \cdot \left(\frac{1}{\rho_{C,j}} - f_L\right)} \tag{59}$$

$$\left.\frac{\partial \mathbf{x}'_{V,j}}{\partial \boldsymbol{\xi}_i}\right|_{(2,2)} = \frac{f_L}{s_y \cdot \left(\frac{1}{\rho_{C,j}} - f_L\right)} \tag{60}$$

$$\left.\frac{\partial \mathbf{x}'_{V,j}}{\partial \boldsymbol{\xi}_i}\right|_{(3,3)} = -\frac{f_L^2}{B \cdot \left(\frac{1}{\rho_{C,j}} - f_L\right)^2} \tag{61}$$

$$\begin{aligned}
\left.\frac{\partial \mathbf{x}'_{V,j}}{\partial \boldsymbol{\xi}_i}\right|_{(1,3)} &= -x_{C,j} \cdot \frac{f_L}{s_x \cdot \left(\frac{1}{\rho_{C,j}} - f_L\right)^2} = -\frac{u'}{\rho_{C,j}} \cdot \frac{f_L}{s_x \cdot \left(\frac{1}{\rho_{C,j}} - f_L\right)^2} \\
&= -u' \cdot \frac{f_L}{\rho_{C,j} \cdot s_x \cdot \left(\frac{1}{\rho_{C,j}^2} - 2 \cdot \frac{f_L}{\rho_{C,j}} + f_L^2\right)} \\
&= -u' \cdot \frac{f_L \cdot s_x^{-1}}{\frac{1}{\rho_{C,j}} - 2 \cdot f_L + \rho_{C,j} \cdot f_L^2}
\end{aligned} \tag{62}$$

$$\begin{aligned}\left.\frac{\partial \mathbf{x}'_{V,j}}{\partial \boldsymbol{\xi}_i}\right|_{(2,3)} &= -y_{C,j} \cdot \frac{f_L}{s_y \cdot \left(\frac{1}{\rho_{C,j}} - f_L\right)^2} = -\frac{v'}{\rho_{C,j}} \cdot \frac{f_L}{s_y \cdot \left(\frac{1}{\rho_{C,j}} - f_L\right)^2} \\ &= -v' \cdot \frac{f_L}{\rho_{C,j} \cdot s_y \cdot \left(\frac{1}{\rho_{C,j}^2} - 2 \cdot \frac{f_L}{\rho_{C,j}} + f_L^2\right)} \\ &= -v' \cdot \frac{f_L \cdot s_y^{-1}}{\frac{1}{\rho_{C,j}} - 2 \cdot f_L + \rho_{C,j} \cdot f_L^2}\end{aligned} \tag{63}$$

$$\begin{aligned}\left.\frac{\partial \mathbf{x}'_{V,j}}{\partial \boldsymbol{\xi}_i}\right|_{(1,4)} &= -x_{C,j} \cdot y_{C,j} \cdot \frac{f_L}{s_x \cdot \left(\frac{1}{\rho_{C,j}} - f_L\right)^2} = -\frac{u'}{\rho_{C,j}} \cdot \frac{v'}{\rho_{C,j}} \cdot \frac{f_L}{s_x \cdot \left(\frac{1}{\rho_{C,j}} - f_L\right)^2} \\ &= -u' \cdot v' \cdot \frac{f_L \cdot s_x^{-1}}{\rho_{C,j}^2 \cdot \left(\frac{1}{\rho_{C,j}^2} - 2 \cdot \frac{f_L}{\rho_{C,j}} + f_L^2\right)} \\ &= -u' \cdot v' \cdot \frac{f_L \cdot s_x^{-1}}{1 - 2 \cdot \rho_{C,j} \cdot f_L + \rho_{C,j}^2 \cdot f_L^2}\end{aligned} \tag{64}$$

$$\begin{aligned}\left.\frac{\partial \mathbf{x}'_{V,j}}{\partial \boldsymbol{\xi}_i}\right|_{(2,4)} &= -z_{C,j} \cdot \frac{f_L}{s_y \cdot \left(\frac{1}{\rho_{C,j}} - f_L\right)} - y_{C,j}^2 \cdot \frac{f_L}{s_y \cdot \left(\frac{1}{\rho_{C,j}} - f_L\right)^2} \\ &= -\frac{1}{\rho_{C,j}} \cdot \frac{f_L}{s_y \cdot \left(\frac{1}{\rho_{C,j}} - f_L\right)} - \frac{v'^2}{\rho_{C,j}^2} \cdot \frac{f_L}{s_y \cdot \left(\frac{1}{\rho_{C,j}} - f_L\right)^2} \\ &= -\frac{f_L \cdot s_y^{-1}}{1 - \rho_{C,j} \cdot f_L} - v'^2 \cdot \frac{f_L \cdot s_y^{-1}}{1 - 2 \cdot \rho_{C,j} \cdot f_L + \rho_{C,j}^2 \cdot f_L^2}\end{aligned} \tag{65}$$

$$\begin{aligned}\left.\frac{\partial \mathbf{x}'_{V,j}}{\partial \boldsymbol{\xi}_i}\right|_{(3,4)} &= -y_{C,j} \cdot \frac{f_L^2}{B \cdot \left(\frac{1}{\rho_{C,j}} - f_L\right)^2} = -\frac{v'}{\rho_{C,j}} \cdot \frac{f_L^2}{B \cdot \left(\frac{1}{\rho_{C,j}} - f_L\right)^2} \\ &= -v' \cdot \frac{f_L}{\rho_{C,j} \cdot B \cdot \left(\frac{1}{\rho_{C,j}^2} - 2 \cdot \frac{f_L}{\rho_{C,j}} + f_L^2\right)} \\ &= -v' \cdot \frac{f_L \cdot B^{-1}}{\frac{1}{\rho_{C,j}} - 2 \cdot f_L + \rho_{C,j} \cdot f_L^2}\end{aligned} \tag{66}$$

$$\left.\frac{\partial \mathbf{x}'_{V,j}}{\partial \boldsymbol{\xi}_i}\right|_{(1,5)} = z_{C,j} \cdot \frac{f_L}{s_x \cdot \left(\frac{1}{\rho_{C,j}} - f_L\right)} + x_{C,j}^2 \cdot \frac{f_L}{s_x \cdot \left(\frac{1}{\rho_{C,j}} - f_L\right)^2}$$
$$= \frac{1}{\rho_{C,j}} \cdot \frac{f_L}{s_x \cdot \left(\frac{1}{\rho_{C,j}} - f_L\right)} + \frac{u'^2}{\rho_{C,j}^2} \cdot \frac{f_L}{s_x \cdot \left(\frac{1}{\rho_{C,j}} - f_L\right)^2} \tag{67}$$
$$= \frac{f_L \cdot s_x^{-1}}{1 - \rho_{C,j} \cdot f_L} + u'^2 \cdot \frac{f_L \cdot s_x^{-1}}{1 - 2 \cdot \rho_{C,j} \cdot f_L + \rho_{C,j}^2 \cdot f_L^2}$$

$$\left.\frac{\partial \mathbf{x}'_{V,j}}{\partial \boldsymbol{\xi}_i}\right|_{(2,5)} = y_{C,j} \cdot x_{C,j} \cdot \frac{f_L}{s_y \cdot \left(\frac{1}{\rho_{C,j}} - f_L\right)^2} = \frac{v'}{\rho_{C,j}} \cdot \frac{u'}{\rho_{C,j}} \cdot \frac{f_L}{s_y \cdot \left(\frac{1}{\rho_{C,j}} - f_L\right)^2}$$
$$= v' \cdot u' \cdot \frac{f_L \cdot s_y^{-1}}{\rho_{C,j}^2 \cdot \left(\frac{1}{\rho_{C,j}^2} - 2 \cdot \frac{f_L}{\rho_{C,j}} + f_L^2\right)}$$
$$= v' \cdot u' \cdot \frac{f_L \cdot s_y^{-1}}{1 - 2 \cdot \rho_{C,j} \cdot f_L + \rho_{C,j}^2 \cdot f_L^2} \tag{68}$$

$$\left.\frac{\partial \mathbf{x}'_{V,j}}{\partial \boldsymbol{\xi}_i}\right|_{(3,5)} = x_{C,j} \cdot \frac{f_L^2}{B \cdot \left(\frac{1}{\rho_{C,j}} - f_L\right)^2} = \frac{u'}{\rho_{C,j}} \cdot \frac{f_L}{B \cdot \left(\frac{1}{\rho_{C,j}} - f_L\right)^2}$$
$$= u' \cdot \frac{f_L^2}{\rho_{C,j} \cdot B \cdot \left(\frac{1}{\rho_{C,j}^2} - 2 \cdot \frac{f_L}{\rho_{C,j}} + f_L^2\right)} \tag{69}$$
$$= u' \cdot \frac{f_L^2 \cdot B^{-1}}{\frac{1}{\rho_{C,j}} - 2 \cdot f_L + \rho_{C,j} \cdot f_L^2}$$

$$\left.\frac{\partial \mathbf{x}'_{V,j}}{\partial \boldsymbol{\xi}_i}\right|_{(1,6)} = -y_{C,j} \cdot \frac{f_L}{s_x \cdot \left(\frac{1}{\rho_{C,j}} - f_L\right)} = -\frac{v'}{\rho_{C,j}} \cdot \frac{f_L}{s_x \cdot \left(\frac{1}{\rho_{C,j}} - f_L\right)}$$
$$= -v' \cdot \frac{f_L \cdot s_x^{-1}}{1 - \rho_{C,j} \cdot f_L} \tag{70}$$

$$\left.\frac{\partial \mathbf{x}'_{V,j}}{\partial \boldsymbol{\xi}_i}\right|_{(2,6)} = x_{C,j} \cdot \frac{f_L}{s_y \cdot \left(\frac{1}{\rho_{C,j}} - f_L\right)} = \frac{u'}{\rho_{C,j}} \cdot \frac{f_L}{s_y \cdot \left(\frac{1}{\rho_{C,j}} - f_L\right)}$$
$$= u' \cdot \frac{f_L \cdot s_y^{-1}}{1 - \rho_{C,j} \cdot f_L} \tag{71}$$

After substitution and simplification, the final Jacobian takes the compact form

$$\frac{\partial \mathbf{x}'_{V,j}}{\partial \boldsymbol{\xi}_i} = \begin{bmatrix} \alpha_x & 0 & -u'\alpha_x & -u'v'\alpha_x & (1+u'^2)\alpha_x & -v'\alpha_x \\ 0 & \alpha_y & -v'\alpha_y & -(1+v'^2)\alpha_y & u'v'\alpha_y & u'\alpha_y \\ 0 & 0 & \beta & -v'\beta & u'\beta & 0 \end{bmatrix}, \tag{72}$$

where the scalar factors are

$$\alpha_x = \frac{f_L}{s_x\left(\frac{1}{d_{p2}} - f_L\right)}, \tag{73}$$

$$\alpha_y = \frac{f_L}{s_y\left(\frac{1}{d_{p2}} - f_L\right)}, \tag{74}$$

$$\beta = -\frac{f_L^2}{B\left(\frac{1}{d_{p2}} - f_L\right)^2}. \tag{75}$$

Pose Parameter Derivatives of the Depth Residual. Since the depth residual is defined in a single frame, it does not affect the camera pose. Thus, the derivative components of the depth residual with respect to the pose are zero (see Eq. (76)).

$$\frac{\partial r^{\text{depth}}}{\partial v_x} = \frac{\partial r^{\text{depth}}}{\partial v_y} = \frac{\partial r^{\text{depth}}}{\partial v_z} = \frac{\partial r^{\text{depth}}}{\partial w_x} = \frac{\partial r^{\text{depth}}}{\partial w_y} = \frac{\partial r^{\text{depth}}}{\partial w_z} = 0 \tag{76}$$

E.3 Inverse Depth Derivative

Inverse Depth Derivatives of the Photometric Residual. The derivative of the photometric residual can be calculated as a function of the inverse metric depth $\rho_{C,i}$ of the image point by decomposing the derivative using the chain rule. It can therefore be expressed as follows:

$$\frac{\partial r^{\text{photo}}}{\partial \rho_{C,i}} = \frac{\partial r^{\text{photo}}}{\partial \mathbf{x}'_{V,j}} \cdot \frac{\partial \mathbf{x}'_{V,j}}{\partial \mathbf{x}_{V,j}} \cdot \frac{\partial \mathbf{x}_{V,j}}{\partial \mathbf{x}_{C,j}} \cdot \frac{\partial \mathbf{x}_{C,j}}{\partial \rho_{C,i}}. \tag{77}$$

Most of the components of Eq. (77) have already been calculated. They can be found in Eq. (29) (image gradient at a given point), Eq. (35), and Eq. (46). Furthermore, the derivative of the point in the dimensionless virtual domain as a function of the inverse metric depth in the camera coordinate system has also already been determined in Eq. (47). It remains to compute $\frac{\partial \mathbf{x}_{C,j}}{\partial \rho_{C,i}}$.

The mathematical transformation to move from the point $\mathbf{x}_{C,i}$ in camera coordinate system of frame I_i to the point $\mathbf{x}_{C,j}$ in camera coordinate system of frame I_j is expressed in Eq. (15). We also use the relationship demonstrated in Eq. (17). We define the components of $\mathbf{R}, \mathbf{t} \in SE(3)$ as follows:

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}, \qquad \mathbf{t} = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}. \tag{78}$$

As before, we define $z_{C,i} = \frac{1}{\rho_{C,i}}$. Combining Eq. (15), Eq. (17), and Eq. (78) yields the following relationship:

$$\begin{aligned} \overline{\mathbf{x}}_{C,j} &= \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \cdot \begin{bmatrix} \frac{z_{C,i}}{b_{L,i}} \left(x'_{V,i} - c_x \right) s_x \\ \frac{z_{C,i}}{b_{L,i}} \left(y'_{V,i} - c_y \right) s_y \\ \left(\frac{1}{f_L} - \frac{1}{b_{L,i}} \right)^{-1} \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \\ \Leftrightarrow \qquad \overline{\mathbf{x}}_{C,j} &= \begin{bmatrix} x_{C,j} \\ y_{C,j} \\ z_{C,j} \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{b_{L,i} \cdot \rho_{C,i}} \left(x'_{V,i} - c_x \right) s_x \\ \frac{1}{b_{L,i} \cdot \rho_{C,i}} \left(y'_{V,i} - c_y \right) s_y \\ \frac{1}{\rho_{C,i}} \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \\ \Leftrightarrow \qquad \begin{bmatrix} x_{C,j} \\ y_{C,j} \\ z_{C,j} \end{bmatrix} &= \begin{bmatrix} r_{11} \frac{s_x}{b_{L,i}\rho_{C,i}} \left(x'_{V,i} - c_x \right) + r_{12} \frac{s_y}{b_{L,i}\rho_{C,i}} \left(y'_{V,i} - c_y \right) + r_{13} \frac{1}{\rho_{C,i}} + t_x \\ r_{21} \frac{s_x}{b_{L,i}\rho_{C,i}} \left(x'_{V,i} - c_x \right) + r_{22} \frac{s_y}{b_{L,i}\rho_{C,i}} \left(y'_{V,i} - c_y \right) + r_{23} \frac{1}{\rho_{C,i}} + t_y \\ r_{31} \frac{s_x}{b_{L,i}\rho_{C,i}} \left(x'_{V,i} - c_x \right) + r_{32} \frac{s_y}{b_{L,i}\rho_{C,i}} \left(y'_{V,i} - c_y \right) + r_{33} \frac{1}{\rho_{C,i}} + t_z \end{bmatrix}. \end{aligned} \tag{79}$$

– Calculation of $\frac{\partial \mathbf{x}_{C,j}}{\partial \rho_{C,i}}$

From Eq. (79), we can now calculate the derivative $\frac{\partial \mathbf{x}_{C,j}}{\partial \rho_{C,i}}$ (see Eq. (80)). We use also again the substitutions in Eq. (55), Eq. (56) and Eq. (57).

$$\frac{\partial \mathbf{x}_{C,j}}{\partial \rho_{C,i}} = \begin{bmatrix} \frac{\partial x_{C,j}}{\partial \rho_{C,i}} \\ \frac{\partial y_{C,j}}{\partial \rho_{C,i}} \\ \frac{\partial z_{C,j}}{\partial \rho_{C,i}} \end{bmatrix} \tag{80}$$

The different components of Eq. (80) are then calculated independently below.

$$\begin{aligned} \frac{\partial x_{C,j}}{\partial \rho_{C,i}} &= \frac{\partial}{\partial \rho_{C,i}} \left(r_{11} \frac{s_x}{b_{L,i}\rho_{C,i}} \left(x'_{V,i} - c_x \right) + r_{12} \frac{s_y}{b_{L,i}\rho_{C,i}} \left(y'_{V,i} - c_y \right) + r_{13} \frac{1}{\rho_{C,i}} + t_x \right) \\ &= -\frac{1}{\rho_{C,i}^2} \cdot \left(r_{11} \cdot \frac{s_x}{f_L} \cdot \left(x'_{V,i} - c_x \right) + r_{12} \cdot \frac{s_y}{f_L} \cdot \left(y'_{V,i} - c_y \right) + r_{13} \right) \end{aligned} \tag{81}$$

$$
\begin{aligned}
\frac{\partial y_{C,j}}{\partial \rho_{C,i}} &= \frac{\partial}{\partial \rho_{C,i}} \left(r_{21} \frac{s_x}{b_{L,i}\rho_{C,i}} \left(x'_{V,i} - c_x\right) + r_{22} \frac{s_y}{b_{L,i}\rho_{C,i}} \left(y'_{V,i} - c_y\right) + r_{23} \frac{1}{\rho_{C,i}} + t_y \right) \\
&= -\frac{1}{\rho_{C,i}^2} \cdot \left(r_{21} \cdot \frac{s_x}{f_L} \cdot \left(x'_{V,i} - c_x\right) + r_{22} \cdot \frac{s_y}{f_L} \cdot \left(y'_{V,i} - c_y\right) + r_{23} \right)
\end{aligned}
\tag{82}
$$

$$
\begin{aligned}
\frac{\partial z_{C,j}}{\partial \rho_{C,i}} &= \frac{\partial}{\partial \rho_{C,i}} \left(r_{31} \frac{s_x}{b_{L,i}\rho_{C,i}} \left(x'_{V,i} - c_x\right) + r_{32} \frac{s_y}{b_{L,i}\rho_{C,i}} \left(y'_{V,i} - c_y\right) + r_{33} \frac{1}{\rho_{C,i}} + t_z \right) \\
&= -\frac{1}{\rho_{C,i}^2} \cdot \left(r_{31} \cdot \frac{s_x}{f_L} \cdot \left(x'_{V,i} - c_x\right) + r_{32} \cdot \frac{s_y}{f_L} \cdot \left(y'_{V,i} - c_y\right) + r_{33} \right)
\end{aligned}
\tag{83}
$$

– Calculation of $\frac{\partial \mathbf{x}'_{V,j}}{\partial \rho_{C,i}}$

Using the chain rule, we can now calculate $\frac{\partial \mathbf{x}'_{V,j}}{\partial \rho_{C,i}}$ from Eq. (47) and Eq. (80):

$$
\frac{\partial \mathbf{x}'_{V,j}}{\partial \rho_{C,i}} = \frac{\partial \mathbf{x}'_{V,j}}{\partial \mathbf{x}_{C,j}} \cdot \frac{\partial \mathbf{x}_{C,j}}{\partial \rho_{C,i}}. \tag{84}
$$

Inverse Depth Derivatives of the Depth Residual. To calculate the derivative of the inverse virtual depth $\rho_{v,i}$ as a function of the inverse metric depth $\rho_{C,i}$ of the image point, we must first express the relationship between these two quantities. Eq. (2) provides an initial expression for the inverse metric depth $\rho_{C,i}$. This gives the relationship Eq. (85).

$$
\rho_{C,i} = \frac{1}{z_C} = \frac{1}{f_L} - \frac{1}{b_L} = b + b_{L0} = \frac{1}{f_L} - \frac{1}{v_i \cdot B + b_{L0}} = \frac{1}{f_L} - \frac{1}{\frac{1}{\rho_{v,i}} \cdot B + b_{L0}} \tag{85}
$$

In the Eq. (85), we isolate $\rho_{v,i}$, the inverse virtual depth of the point, to get Eq. (86).

$$
\begin{aligned}
\rho_{C,i} &= \frac{1}{f_L} - \frac{1}{\frac{1}{\rho_{v,i}} \cdot B + b_{L0}} \\
\rho_{v,i} &= \frac{1}{\frac{1}{B} \cdot \left(\frac{1}{\frac{1}{f_L} - \rho_{C,i}} - b_{L0} \right)} \\
\rho_{v,i} &= \frac{B - B \cdot \rho_{C,i} \cdot f_L}{\rho_{C,i} \cdot f_L \cdot b_{L0} + f_L - b_{L0}}
\end{aligned}
\tag{86}
$$

Eq. (86) is the expression for the inverse virtual depth estimated by the pipeline. We can now calculate the derivative of r^{depth} using the expression in Eq. (22). The measured value ρ_v^{mes} do not depend on $\rho_{C,i}$. So, the result is given in Eq. (87).

$$\begin{aligned}\frac{\partial r^{\text{depth}}}{\partial \rho_{C,i}} &= \frac{\partial}{\partial \rho_{C,i}} \frac{B - B \cdot \rho_{C,i} \cdot f_L}{\rho_{C,i} \cdot f_L \cdot b_{L0} + f_L - b_{L0}} \\ &= \frac{-B \cdot f_L \cdot (\rho_{C,i} \cdot b_{L0} + f_L - b_{L0}) - (B - B \cdot \rho_{C,i} \cdot f_L) \cdot f_L \cdot b_{L0}}{(\rho_{C,i} \cdot f_L \cdot b_{L0} + f_L - b_{L0})^2} \\ &= -\frac{B \cdot f_L^2}{(f_L \cdot b_{L0} \cdot \rho_{C,i} + f_L - b_{L0})^2}\end{aligned} \tag{87}$$

F RGB Image Generation for Monocular Camera Pipelines

We evaluate our pipeline on the dataset LiFMCR introduced in [4], which contains only plenoptic data. In order to compare our approach with existing monocular pipelines, we generate images compatible with a pinhole camera model following a procedure similar to the one described in [3].

In a plenoptic representation, virtual points $\mathbf{x}'_{V,i} = [x'_{V,i}, y'_{V,i}, z'_{V,i}]^T$ are formed at a distance from the MLA that depends on the depth of the observed scene point $\mathbf{x}_C = [x_C, y_C, z_C]^T$ relative to the camera. As a consequence, these virtual points do not lie on a common image plane, which contradicts the assumptions of the pinhole camera model where all image points are projected onto a single plane.

To address this discrepancy, instead of horizontally projecting the plenoptic virtual points to generate the totally focused image, we reproject them along rays passing through the center of the main lens onto a common projection plane. This operation enforces a central perspective projection consistent with a pinhole camera model. The projection principle is illustrated in Fig. 3.

The projection plane is placed at a distance $2B$ from the MLA. This distance corresponds to the maximum measurable depth while preserving the largest possible depth of field. Projecting all virtual points onto this plane therefore produces image coordinates consistent with a conventional perspective camera model.

Given a virtual point $\mathbf{x}'_V = [x'_V, y'_V, z'_V]^T$, the coordinates of the projected point on the common plane $\mathbf{x}_p = [x_p, y_p]^T$ are obtained by intersecting the ray connecting the virtual point to the center of the main lens with the projection plane. The resulting coordinates are computed using Eq. (88) and Eq. (89).

$$x_p = \frac{x'_V - c_x}{v \cdot B + b_{L0}} \cdot (2 \cdot B + b_{L0}) + c_x \tag{88}$$

$$y_p = \frac{y'_V - c_y}{v \cdot B + b_{L0}} \cdot (2 \cdot B + b_{L0}) + c_y \tag{89}$$

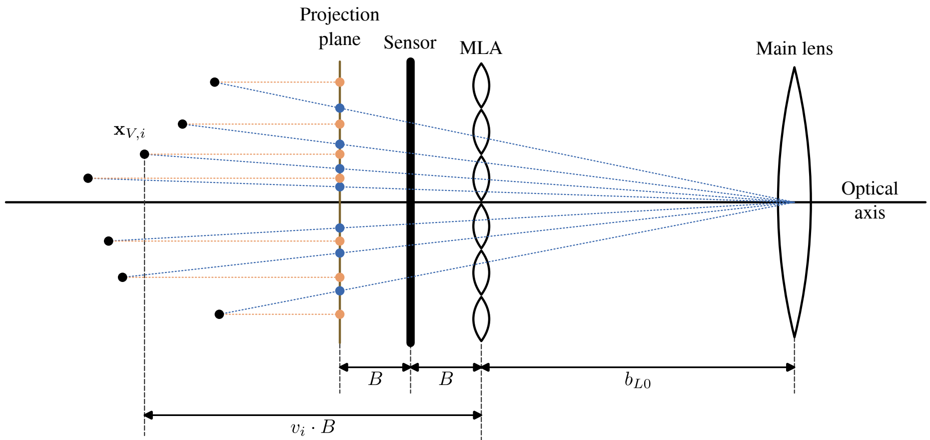

Fig. 3: Projection of the virtual image points onto an image plane at a distance $2B$ from the MLA (inspired by [3]). The points are projected from a central perspective projection through the center of the main lens (in *blue*) instead of a horizontal projection (in *orange*) in order to generate data according to a pinhole model compatible with mono camera pipelines.

G RGB-D Comparison

Our primary comparison focuses on monocular visual odometry (VO)/SLAM, as PRISM-VO operates using a single passive imaging sensor, namely a plenoptic camera. In contrast, RGB-D methods rely on active depth sensing and therefore do not constitute a strictly equivalent comparison. Nevertheless, RGB-D systems provide a valuable reference point and serve as relevant baseline methods.

To this end, we evaluated ORB-SLAM3 [2] in RGB-D mode on the reference dataset [6], using depth maps generated from the synchronized stereo image pairs provided in the dataset. The field of view of the virtual RGB-D camera was adjusted to closely match that of the plenoptic camera while preserving the validity of the pinhole camera model. The resulting cumulative error plots for this evaluation, together with those of PRISM-VO, are shown in Fig. 4. ORB-SLAM3 in RGB-D mode achieves slightly better scale accuracy, which can be attributed to the larger stereo baseline available for depth estimation compared to the much smaller micro-image baselines of a plenoptic camera. However, PRISM-VO demonstrates greater robustness, successfully processing a larger number of sequences.

H Monocular Comparison with Original Field of View

The performance of monocular systems depends on the available field of view. In our evaluations, we aimed to compare under identical sensor observations, as the plenoptic setup is limited in field of view (expandable via other lenses). To

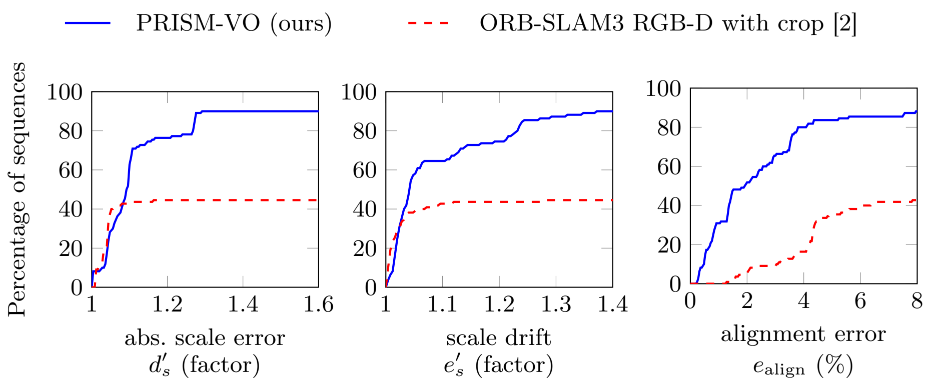

Fig. 4: Cumulative plots obtained on the dataset [6] for PRISM-VO and ORB-SLAM3 in RGB-D mode with cropped images to match the field of view of the plenoptic camera.

(a) Plenoptic focus image **(b)** Monocular original image **(c)** Monocular cropped image

Fig. 5: Field of view comparison between the plenoptic camera and the stereo cameras used in the dataset [6]. In the experiments where this is specified, the field of view of the monocular camera is cropped as shown in Fig. 5c (without violating the plenoptic model).

ensure a fair comparison based on equivalent sensor observations, we matched the field of view of the monocular baseline to that of the plenoptic camera by cropping the images. Note that this procedure preserves the validity of the pinhole camera model through the corresponding adjustment of the intrinsic calibration parameters.

Fig. 5 illustrates the field of view difference between the plenoptic camera and the monocular camera of the reference dataset [6], together with an example of the cropped monocular image used in our evaluation. For completeness, Fig. 6 additionally reports ORB-SLAM3 results obtained in monocular mode using the original, uncropped field of view, since it is consitently failing on the cropped images. Despite operating with a narrower field of view, PRISM-VO achieves slightly lower scale drift, comparable alignment accuracy, and greater robustness across the evaluated sequences of the reference dataset [6].

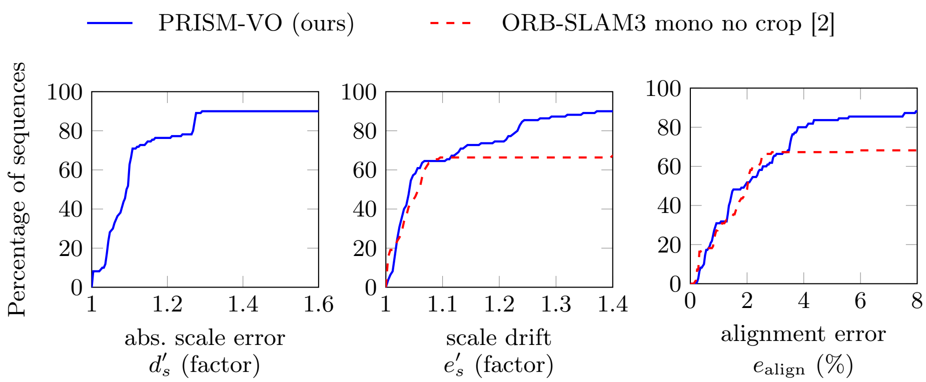

Fig. 6: Cumulative plots obtained on the dataset [6] for PRISM-VO and ORB-SLAM3 in monocular mode with full size images.

References

1. Brown, D.: Decentering distortion of lenses. Photogrammetric Engineering **32**(3), 444–462 (1966)
2. Campos, C., Elvira, R., Rodríguez, J.J.G., Montiel, J.M., Tardós, J.D.: ORB-SLAM3: An accurate open-source library for visual, visual-inertial, and multi-map SLAM. Transactions on Robotics (T-RO) **37**(6), 1874–1890 (2021). `https://doi.org/10.1109/TRO.2021.3075644`
3. Fleith, A., Ahmed, D., Cremers, D., Zeller, N.: LiFCal: Online light field camera calibration via bundle adjustment. In: DAGM German Conference on Pattern Recognition (GCPR). pp. 120–136. Springer (2024). `https://doi.org/10.1007/978-3-031-85187-2_8`
4. Fleith, A., Zirbel, J., Cremers, D., Zeller, N.: LiFMCR: Dataset and benchmark for light field multi-camera registration. In: International Symposium on Visual Computing (ISVC). Springer Nature Switzerland (2026). `https://doi.org/10.1007/978-3-032-14492-8_35`
5. Perwass, C., Wietzke, L.: Single-lens 3D camera with extended depth of field. In: Human Vision and Electronic Imaging (HVEI). vol. 8291, pp. 45–59. SPIE (2012). `https://doi.org/10.1117/12.909882`
6. Zeller, N., Quint, F., Stilla, U.: A synchronized stereo and plenoptic visual odometry dataset. arXiv preprint arXiv:1807.09372 (2018). `https://doi.org/10.48550/arXiv.1807.09372`