Camera Extrinsics Calibration: Theory

Mathematical formulation of the residuals used in the optimization.

May 7, 2019

1 Reference Frames

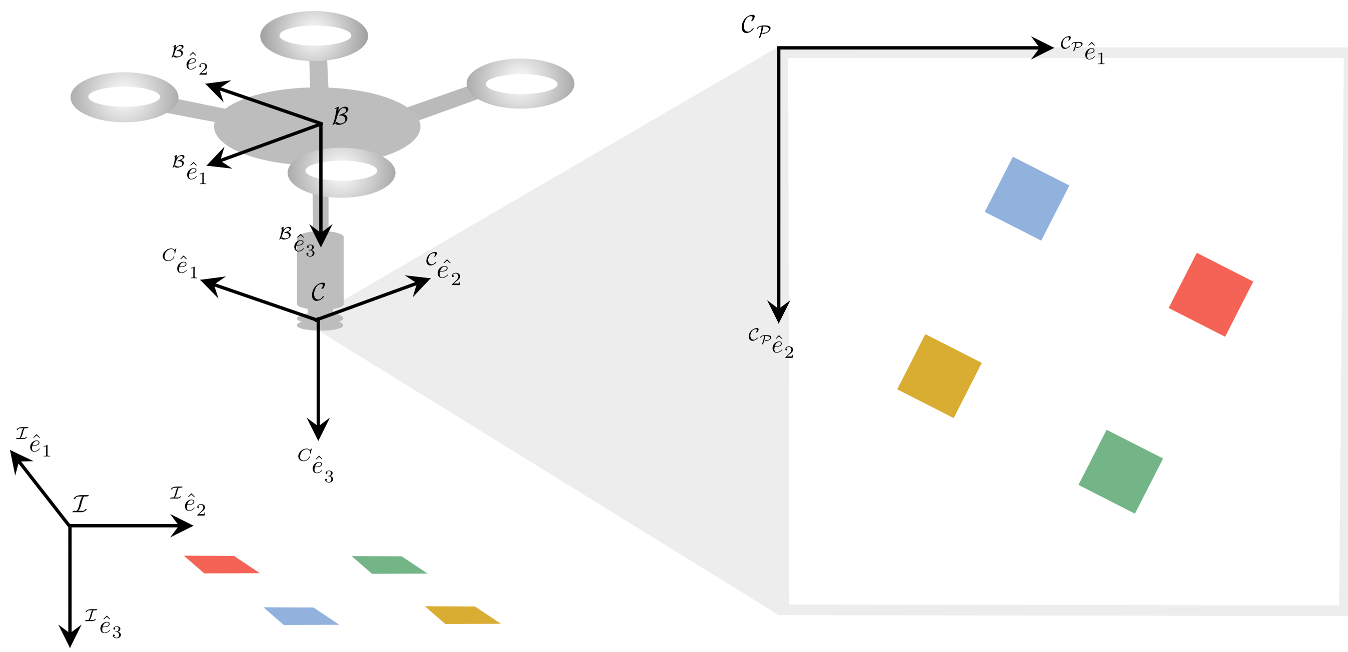

Figure 1: Reference frames for the camera extrinsics calibration problem.

All vector quantities in the camera extrinsics calibration problem are expressed in one of the following frames:

- **The inertial/global frame ($\mathcal{I}$):** This is the north-east-down (NED) frame, assumed to be stationary over the course of the calibration routine.
- **The UAV body frame ($\mathcal{B}$):** This frame is body-centric and body-fixed to the UAV. The x-axis sticks out the front of the UAV, the y-axis out the right side, and the z-axis out the bottom, towards the ground.
- **The camera frame ($\mathcal{C}$):** This frame is centered on the pinhole convergence point of the camera attached to the UAV. From the perspective of the camera image, the x-axis sticks out to the right, the y-axis points down, and the z-axis points out of the camera plane toward the world.
- **The camera pixel frame ($\mathcal{C}_{\mathcal{P}}$):** This frame coincides with the camera frame in terms of orientation, but is centered on the image plane, where units are measured in pixels.

2 Definitions of Relevant $\mathbb{R}^n$ and $\mathbb{S}^3$ Quantities

The following vectors in $\mathbb{R}^n$ and manifold objects in $\mathbb{S}^3$ (the space of unit-length quaternions*) are relevant to the calculations in this derivation:

Quantity	Explanation
${}^{\mathcal{B}}x_{\mathcal{C},0} \in \mathbb{R}^3$	The **initial guess** for the translational offset between the origin of the UAV body frame and the origin of the camera frame. Expressed in the body frame.
$q^{\mathcal{C}}_{\mathcal{B},0} \in \mathbb{S}^3$	The **initial guess** for the (passive) rotational offset between the UAV body frame and the camera frame.
${}^{\mathcal{B}}x_{\mathcal{C},f} \in \mathbb{R}^3$	The **calculated** translational offset between the origin of the UAV body frame and the origin of the camera frame, as a result of the optimization routine. Expressed in the body frame.
$q^{\mathcal{C}}_{\mathcal{B},f} \in \mathbb{S}^3$	The **calculated** (passive) rotational offset between the UAV body frame and the camera frame, as a result of the optimization routine.
${}^{\mathcal{I}}x_{\mathcal{B}} \in \mathbb{R}^3$	The translation vector of the UAV with respect to the inertial origin at the current time step. Expressed in the inertial frame.
$q^{\mathcal{B}}_{\mathcal{I}} \in \mathbb{S}^3$	The (passive) rotation of the UAV with respect to the inertial frame at the current time step.
${}^{\mathcal{I}}l_i \in \mathbb{R}^3$	The position of stationary visual landmark i with respect to the inertial origin. Expressed in the inertial frame.
${}^{\mathcal{C}_{\mathcal{P}}}p_i \in \mathbb{R}^2$	The measured pixel coordinates $\begin{bmatrix} u_i & v_i \end{bmatrix}^T$ of landmark i in the image plane. Expressed in the camera pixel frame.
${}^{\mathcal{C}_{\mathcal{P}}}\hat{p}_i \in \mathbb{R}^2$	The **theoretical** pixel coordinates $\begin{bmatrix} \hat{u}_i & \hat{v}_i \end{bmatrix}^T$ of landmark i in the image plane, given the initial guess for the extrinsic camera parameters ${}^{\mathcal{B}}x_{\mathcal{C},0}$ and $q^{\mathcal{C}}_{\mathcal{B},0}$.
$f \in \mathbb{R}^2$	The focal lengths $\begin{bmatrix} f_x & f_y \end{bmatrix}^T$ of the camera.
$c \in \mathbb{R}^2$	The center pixels $\begin{bmatrix} c_x & c_y \end{bmatrix}^T$ of the camera's image plane.
$s \in \mathbb{R}^1$	The skew of the camera, which defines horizontal shear of the pixels in the image plane.
$d \in \mathbb{R}^5$	The camera's (radial) distortion parameters $\begin{bmatrix} k_1 & k_2 & p_1 & p_2 & k_3 \end{bmatrix}^T$; see the radial distortion model definition in Section 3.

* The software library used in this optimization routine implements quaternion math operations, hence the reference to the manifold objects over $\mathbb{S}^3$. However, for concision and clarity, the following sections detail the underlying mathematics using rotation matrices instead of quaternions:

$$q^b_a \in \mathbb{S}^3 \longleftrightarrow R^b_a \in SO(3)$$

Providing the transformation operations in terms of rotation matrices gives an accurate (albeit less computationally efficient) representation of the underlying mathematics in the optimization routine.

3 Camera Model

The following camera models assume a pinhole camera model, accounting for radial distortion and skew.

3.1 Transforming to the Camera Frame

Given an inertial landmark position ${}^{\mathcal{I}}l_i$, the landmark coordinates are transformed into the camera frame, simultaneously translating the origin to coincide with the origin of the camera frame using rigid body homogeneous transform matrices $H_a^b \in SE(3)$. This requires projecting ${}^{\mathcal{I}}l_i$ into homogeneous coordinates ${}^{\mathcal{I}}\mathbf{l}_i \in \mathbb{R}^4$:

$${}^{\mathcal{I}}\mathbf{l}_i = \begin{bmatrix} {}^{\mathcal{I}}l_i & 1 \end{bmatrix}^T$$

$$H_{\mathcal{I}}^{\mathcal{B}} = \begin{bmatrix} R_{\mathcal{I}}^{\mathcal{B}} & R_{\mathcal{I}}^{\mathcal{B}\mathcal{I}}x_{\mathcal{B}} \\ 0 & 1 \end{bmatrix}$$

$$H_{\mathcal{B}}^{\mathcal{C}} = \begin{bmatrix} R_{\mathcal{B}}^{\mathcal{C}} & R_{\mathcal{B}}^{\mathcal{C}\mathcal{B}}x_{\mathcal{C}} \\ 0 & 1 \end{bmatrix}$$

$$H_{\mathcal{I}}^{\mathcal{C}} = H_{\mathcal{B}}^{\mathcal{C}} H_{\mathcal{I}}^{\mathcal{B}}$$

$${}^{\mathcal{C}}\mathbf{l}_i = \begin{bmatrix} l_{x,i} & l_{y,i} & l_{z,i} & 1 \end{bmatrix}^T = H_{\mathcal{I}}^{\mathcal{C}\mathcal{I}}\mathbf{l}_i$$

For convenience in subsequent operations, we project ${}^{\mathcal{C}}\mathbf{l}_i$ back onto $\mathbb{R}^3$ and divide all components by the z-component:

$${}^{\mathcal{C}}l_i = \begin{bmatrix} l_{x,i}/l_{z,i} & l_{y,i}/l_{z,i} & 1 \end{bmatrix}^T$$

3.2 Radial Distortion Model

It is assumed that the UAV camera imposes some kind of radial distortion on measured pixel features. Thus, radial distortion is applied to the point ${}^{\mathcal{C}}l_i$ before it is projected onto the camera pixel plane. Given a *distorted* feature location in the camera frame ${}^{\mathcal{C}}l_{i,d}$, the corresponding *undistorted* feature location in the camera frame ${}^{\mathcal{C}}l_i$ is obtained according to the following radial undistortion model:

$${}^{\mathcal{C}}l_i = f({}^{\mathcal{C}}l_{i,d}, d) = g \begin{bmatrix} {}^{\mathcal{C}}l_{i,d,x} + 2p_1{}^{\mathcal{C}}l_{i,d,x}{}^{\mathcal{C}}l_{i,d,y} + p_2(({}^{\mathcal{C}}l_{i,d,x}{}^{\mathcal{C}}l_{i,d,y})^2 + 2{}^{\mathcal{C}}l_{i,d,x}{}^{\mathcal{C}}l_{i,d,x}) \\ {}^{\mathcal{C}}l_{i,d,y} + 2p_2{}^{\mathcal{C}}l_{i,d,x}{}^{\mathcal{C}}l_{i,d,y} + p_1(({}^{\mathcal{C}}l_{i,d,x}{}^{\mathcal{C}}l_{i,d,y})^2 + 2{}^{\mathcal{C}}l_{i,d,y}{}^{\mathcal{C}}l_{i,d,y}) \end{bmatrix},$$

where $g = 1 + k_1({}^{\mathcal{C}}l_{i,d,x}{}^{\mathcal{C}}l_{i,d,y})^2 + k_2({}^{\mathcal{C}}l_{i,d,x}{}^{\mathcal{C}}l_{i,d,y})^4 + k_3({}^{\mathcal{C}}l_{i,d,x}{}^{\mathcal{C}}l_{i,d,y})^6$.

To apply distortion to ${}^{\mathcal{C}}l_i$ from Section 3.1 to obtain ${}^{\mathcal{C}}l_{i,d}$, the undistortion model $f({}^{\mathcal{C}}l_{i,d}, d)$ must be inverted. The model cannot be inverted explicitly, so it is inverted iteratively using Newton's method of function minimization.

3.3 Pinhole Projection Model

Having obtained the distorted feature location in the camera frame, a pinhole projection model is used to obtain the distorted feature *pixel* location in the camera image plane. The projection model comes from similar triangles and the definition of horizontal shear:

$${}^{\mathcal{C}_{\mathcal{P}}}\hat{p}_i = \begin{bmatrix} u_i \\ v_i \end{bmatrix} = \begin{bmatrix} f_x{}^{\mathcal{C}}l_{i,d,x} + s{}^{\mathcal{C}}l_{i,d,y} + c_x \\ f_y{}^{\mathcal{C}}l_{i,d,y} + c_y \end{bmatrix}$$

4 Residual Calculation

The calibration program uses nonlinear optimization techniques to minimize the following cost function:

$$J = \sum_{t}^{T} \sum_{i}^{N} \left({}^{\mathcal{C}_{\mathcal{P}}} p_{i,t} - {}^{\mathcal{C}_{\mathcal{P}}} \hat{p}_{i,t} \right)^2$$

for landmark feature $i = 1, \ldots, N$ and time step $t = 1, \ldots, T$. All ${}^{\mathcal{C}_{\mathcal{P}}} p_{i,t}$ are given as an input dataset to the program, and the optimizer modifies ${}^{\mathcal{B}} x_{\mathcal{C}}$ and $q_{\mathcal{B}}^{\mathcal{C}}$ (the extrinsic parameters, or offsets), which are used in the calculation of ${}^{\mathcal{C}_{\mathcal{P}}} \hat{p}_{i,t}$, to minimize the error. It is not necessary for a residual to be calculated at every possible (i, t) combination for the optimization to work, though a wide breadth of measurements (and a varied flight path, producing diversity in ${}^{\mathcal{I}} x_{\mathcal{B}}$ and $q_{\mathcal{I}}^{\mathcal{B}}$ at each time step) increases the chances of convergence.