Corner localization and camera calibration from imaged lattices (supplementary material)

Andrii Stadnik[1] and James Pritts[2]

[1]Faculty of Computer Science, Ukrainian Catholic University, Lviv, Ukraine
[2]Czech Institute of Informatics, Robotics and Cybernetics at Czech Technical University in Prague, Czechia

1 Camera model

1.1 Notation

Term	Description
$\mathbf{u} = \left(u, v, 1\right)^T$	A point in the image space
$\mathbf{x} = \left(x, y, z, 1\right)^T$	A point in the world space
R	A 3×3 rotation matrix
$\mathbf{t}$	A translation vector of length 3
α_x, α_y	Scale factor in the x and y direction
c_x, c_y	Coordinates of the principal point
θ	Angle between the x and y pixel axes
f	Distance from the camera center to the image plane (focal length)
f_x, f_y	Focal lengths in the x and y directions
K	Intrinsic matrix
H	A 3×3 matrix viewing $z = 0$
λ_n	Distortion coefficients

Table 1: Notation

1.2 Extrinsic parameters

The extrinsic parameters represent a rigid transformation from a 3-D world coordinate system to the 3-D camera's coordinate system.

$$\hat{\mathbf{x}} = R\left(x, y, z\right)^T + \mathbf{t} = \begin{bmatrix} R & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix}\left(x, y, z, 1\right)^T,$$

where $\left(x, y, z\right)^T$ is a 3D scene point, R is a 3×3 rotation matrix and $\mathbf{t}$ is a 3×1 translation vector.

Assuming coplanar scene points ($z = 0$):

$$\alpha \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{bmatrix} \mathbf{r_1} & \mathbf{r_2} & \mathbf{r_3} & \mathbf{t} \end{bmatrix} \begin{pmatrix} x \\ y \\ 0 \\ 1 \end{pmatrix} = \underbrace{\begin{bmatrix} \mathbf{r_1} & \mathbf{r_2} & \mathbf{t} \end{bmatrix}}_{H} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}.$$

1.3 Distortion model

The distortion of the image is caused by the lens not being perfectly planar. We used the division distortion model (Fitzgibbon, 2001) which maps a point from a retinal plane to the ray direction in the camera coordinates system.

$$g(\hat{\mathbf{u}}) = \left(u, v, \psi(r(\hat{\mathbf{u}}))\right)^T, \psi(\rho) = 1 + \sum_{n=1}^{N} \lambda_n \rho^{2n},$$

where $\hat{\mathbf{u}} = \left(u, v, 1\right)^T$ is a point in the retinal plane , $r(\hat{\mathbf{u}}) = \sqrt{u^2 + v^2}$ is the radial distance from the principal point and λ_n are the distortion coefficients.

1.3.1 Back-projection using the Division Model

The function $\psi(\cdot)$ is not invertible in general. Let $\hat{\mathbf{x}} = \left(x, y, z\right)^T = \alpha g(\hat{\mathbf{u}})$ be a ray in the camera coordinate system. Then,

$$\frac{\mathbf{x}}{z} = \left(\frac{x}{z}, \frac{y}{z}, 1\right)^T = \left(\frac{\alpha u}{\alpha\psi(r(\hat{\mathbf{u}}))}, \frac{\alpha v}{\alpha\psi(r(\hat{\mathbf{u}}))}, 1\right)^T = \left(\frac{u}{\psi(r(\hat{\mathbf{u}}))}, \frac{v}{\psi(r(\hat{\mathbf{u}}))}, 1\right)^T. \tag{1}$$

From 1 we see that

$$\begin{cases} \frac{x}{z} = \frac{u}{\psi(r(\hat{\mathbf{u}}))} \\ \frac{y}{z} = \frac{v}{\psi(r(\hat{\mathbf{u}}))} \end{cases} \implies \begin{cases} u = \frac{x\psi(r(\hat{\mathbf{u}}))}{z} \\ v = \frac{y\psi(r(\hat{\mathbf{u}}))}{z} \end{cases}. \tag{2}$$

Now, let $\hat{r}$ be a root of $r(\hat{\mathbf{u}}) = \sqrt{\frac{x*\psi(\hat{r})}{z}^2 + \frac{y*\psi(\hat{r})}{z}^2} = \hat{r}$, assuming $0 \le \hat{r} \le w$ where w is a width of the image.

Then, $\hat{\mathbf{u}} = f(\mathbf{x}) = \frac{\hat{r}}{r(\mathbf{x})}\mathbf{x}$, where $f(\cdot)$ is the inverse of $g(\cdot)$.

1.4 Intrinsic parameters

Represent a projective transformation from the 3-D camera's coordinates into the 2-D image coordinates.

$$K = \begin{bmatrix} \alpha_x & \alpha_x \cot\theta & c_x \\ 0 & \alpha_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} f_x & k & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}.$$

For a typical camera, $\theta = \pi/2$ and $\alpha_x = \alpha_y$ (Hartley and Zisserman, 2004):

$$K = \begin{bmatrix} f & 0 & c_x \\ 0 & f & c_y \\ 0 & 0 & 1 \end{bmatrix}.$$

1.5 Complete projection and backprojection

$$\alpha\mathbf{u} = Kf(H\mathbf{x}) \quad \text{(Projection)}$$
$$\alpha\mathbf{x} = H^{-1}g(K^{-1}\mathbf{u}). \quad \text{(Back projection)}$$

2 Approach

2.1 Camera calibration

2.1.1 Initial approximation

2.1.2 Solving for the camera extrinsic parameters

To derive the solver for the camera extrinsic parameters, start from eq. (Projection):

$$\alpha \mathbf{u} = K f(H\mathbf{x}) \tag{3}$$

$$\alpha K^{-1}\mathbf{u} = f(H\mathbf{x}) \quad \text{Move } K \text{ to the left side} \tag{4}$$

$$\alpha g(K^{-1}\mathbf{u}) = g(f(H\mathbf{x})) \quad \text{Set } \widehat{\mathbf{u}} = K^{-1}\mathbf{u}\text{; apply } g(\cdot) \tag{5}$$

$$\alpha \begin{pmatrix} \widehat{u}_x \\ \widehat{u}_y \\ \psi(r(\widehat{\mathbf{u}})) \end{pmatrix}^T = H\mathbf{x}. \tag{6}$$

To eliminate the dependency on the scale α, multiply both sides vectorially by $g(\widehat{\mathbf{u}})$:

$$\alpha g(\widehat{\mathbf{u}}) \times g(\widehat{\mathbf{u}}) = g(\widehat{\mathbf{u}}) \times H\mathbf{x} \implies g(\widehat{\mathbf{u}}) \times \begin{bmatrix} \mathbf{r_1} & \mathbf{r_2} & \mathbf{t} \end{bmatrix} \mathbf{x} = 0. \tag{7}$$

From eq. (7), we can see that a point contributes to three homogeneous equations:

$$\widehat{v}(r_{31}x + r_{32}y + t_3) - g(r(\widehat{\mathbf{u}}))(r_{21}x + r_{22}y + t_2) = 0 \tag{8}$$

$$g(r(\widehat{\mathbf{u}}))(r_{11}x + r_{12}y + t_1) - \widehat{u}(r_{31}x + r_{32}y + t_3) = 0 \tag{9}$$

$$\widehat{u}(r_{21}x + r_{22}y + t_2) - \widehat{v}(r_{11}x + r_{12}y + t_1) = 0. \tag{10}$$

Only eq. (10) is linear in the unknowns. Each point gives a single equation. Now, by rewriting the equation in the matrix form $M \cdot \mathbf{h} = 0$, where

$$\mathbf{h} = \left(r_{11}, r_{12}, r_{21}, r_{22}, r_{31}, r_{32}\right)^T,$$

we get:

$$M = \begin{bmatrix} -\widehat{v}_1 x_1 & -\widehat{v}_1 y_1 & -\widehat{u}_1 x_1 & -\widehat{u}_1 y_1 & -\widehat{v}_1 & -\widehat{u}_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -\widehat{v}_N x_N & -\widehat{v}_N y_N & -\widehat{u}_N x_N & -\widehat{u}_N y_N & -\widehat{v}_N & -\widehat{u}_N \end{bmatrix} \tag{11}$$

, where N is the number of points.

The linear estimate of $\mathbf{h}$ is found by minimizing $||M \cdot \mathbf{h}||^2$ using SVD. The solution is known up to a scale factor.

To find r_{31} and r_{33}, note that $\mathbf{r_1}$ and $\mathbf{r_2}$ are orthonormal. The derivation couldn't fit here, you can find it in the thesis paper.

2.1.3 Extrinsic parameters

Now, to find the rest of the parameters, we substitute the values, found in the previous step into eq. (8) and eq. (9). We assumed the number of the division model's parameter to be equal to 2, and the scalar multiplier to be equal to 1 section 1.3.1:

$$\begin{bmatrix} A_1\rho_1^2 & A_1\rho_1^4 & -v_1 \\ C_1\rho_1^2 & C_1\rho_1^4 & -v_1 \\ \vdots & \vdots & \vdots \\ A_N\rho_N^2 & A_N\rho_N^4 & -v_N \\ C_N\rho_N^2 & C_N\rho_N^4 & -v_N \end{bmatrix} \cdot \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ t_3 \end{bmatrix} = \begin{bmatrix} B_1 - A_1 \\ D_1 - C_1 \\ \vdots \\ B_N - A_N \\ D_N - C_N \end{bmatrix}, \tag{12}$$

where

$$A_i = r_{21}x_i + r_{22}y_i + t_2 \tag{13}$$

$$B_i = v_i(r_{31}x_i + r_{32}y_i) \tag{14}$$

$$C_i = r_{11}x_i + r_{12}y_i + t_1 \tag{15}$$

$$D_i = u_i(r_{31}x_i + r_{32}y_i). \tag{16}$$

The solution can be found using the least squares method.