Multiple View Geometry: Exercise Sheet 4

Prof. Dr. Daniel Cremers,
Mohammed Brahimi, Zhenzhang Ye, Regine Hartwig
Computer Vision Group, TU Munich
`Wednesdays 16:15–17:45 at Hörsaal 2, "Interims I" (5620.01.102), and on RBG Live`

Exercise: May 22nd, 2024

Part I: Theory

1. **Image Formation**

We are looking at the formation of an image in camera coordinates $\mathbf{X} = (X\ Y\ Z\ 1)^\top$. In the lecture, you learned the following relation of homogeneous pixel coordinates $\mathbf{x}'$ and $\mathbf{X}$:

$$\lambda \mathbf{x}' = K \Pi_0 \mathbf{X} \tag{1}$$

with the intrinsic camera matrix K. To clearly differentiate between camera coordinates and pixel coordinates, call the pixel coordinates u and v: $\mathbf{x}' = (u\ v\ 1)^\top$. Furthermore, let the non-homogeneous camera coordinates be $\tilde{\mathbf{X}} := \Pi_0 \mathbf{X} = (X\ Y\ Z)^\top$. (1) is then equivalent to

$$\lambda \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = K \tilde{\mathbf{X}}\,. \tag{2}$$

Let $s_x = s_y = 1$ and $s_\theta = 0$ in the intrinsic camera matrix.

(a) Compute λ and show that (2) is equivalent to

$$u = \frac{fX}{Z} + o_x\,, \quad v = \frac{fY}{Z} + o_y\,. \tag{3}$$

(b) A classic ambiguity of the perspective projection is that one cannot tell an object from another object that is exactly *twice as big but twice as far*. Explain why this is true.

(c) For a camera with $f = 540$, $o_x = 320$ and $o_y = 240$, compute the pixel coordinates u and v of a point $\tilde{\mathbf{X}} = (60\ 100\ 180)^\top$. Explain with the help of (b) why the units of $\tilde{\mathbf{X}}$ are not needed for this task. Will the projected point be in the image if it has dimensions 640×480?

We define the generic projection π of $\tilde{\mathbf{X}}$ to 2D coordinates as follows:

$$\pi(\tilde{\mathbf{X}}) := \begin{pmatrix} X/Z \\ Y/Z \end{pmatrix} \tag{4}$$

(d) Using the generic projection π, show that (3) — and therefore also (1) and (2) — is equivalent to

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = K \begin{pmatrix} \pi(\tilde{\mathbf{X}}) \\ 1 \end{pmatrix}\,. \tag{5}$$

2. **Radial Distortion**

A general image formation model for radially distorted cameras is generic projection followed by a non-linear transformation of the radius for each image point. The distorted coordinates of a generically projected point $\tilde{\mathbf{X}}$ are given by

$$\pi_d(\tilde{\mathbf{X}}) = g\left(\|\pi(\tilde{\mathbf{X}})\|\right) \cdot \pi(\tilde{\mathbf{X}}) \in \mathbb{R}^2 \,. \tag{6}$$

$g : \mathbb{R}^+ \to \mathbb{R}^+$ is the function that radially distorts the coordinates of $\pi(\tilde{\mathbf{X}})$. It is typically approximated by some parametric function. The pixel coordinates of the distorted camera are

$$\begin{pmatrix} u_d \\ v_d \\ 1 \end{pmatrix} = K \begin{pmatrix} \pi_d(\tilde{\mathbf{X}}) \\ 1 \end{pmatrix} = K \begin{pmatrix} g\left(\|\pi(\tilde{\mathbf{X}})\|\right) \cdot \pi(\tilde{\mathbf{X}}) \\ 1 \end{pmatrix} \,. \tag{7}$$

(a) Can this model be used for lenses with a field of view of more than $180°$?

The so-called FOV or ATAN model, first suggested by Devernay and Faugeras in 2001, and used e.g. in the open source implementation of PTAM (Parallel Tracking and Mapping), is given by

$$g_{\text{ATAN}}(r) = \frac{1}{\omega r} \arctan\left(2r \tan\left(\frac{\omega}{2}\right)\right) \,. \tag{8}$$

(b) Derive a closed form solution for f in the undistortion formula

$$\pi(\tilde{\mathbf{X}}) = f\left(\|\pi_d(\tilde{\mathbf{X}})\|\right) \cdot \pi_d(\tilde{\mathbf{X}}) \tag{9}$$

using (6) and $g(r) = g_{\text{ATAN}}(r)$.
Hint: compute the norm of both sides of the equation in (6) and (9).

Note (additional information):
Both formulations of a distortion model, (6) and (9), can be found in the literature. (9) is the one that was presented in the lecture for a polynomial f. In order to switch between the two formulations, it is beneficial if $g(r)r$ and $f(r_d)r_d$ are invertible in closed form. For that reason, the ATAN model is very popular. Another popular choice for radial distortion functions are polynomials.