Arriving at Z/f

Michael Horná\v{c}ek
CVPR 2014 Submission #1678
Supplementary Material

We state in Section 2.1 of the paper that it is easy to show that the distance between two points both situated at the same depth Z and projecting to neighboring pixels is given by Z/f, where f is the camera's focal length in units of pixels (cf. Figure 1). We motivate this result by computing the distance between the point $\mathbf{P}$ at depth Z projecting to the pixel $(i, j)^\top$, and the points at the same depth Z projecting to the pixel neighbors $(i+1, j)^\top, (i, j+1)^\top$, respectively. The point $\mathbf{P}$ is given by $Z \cdot \mathtt{K}^{-1}(i, j, 1)^\top$, where $\mathtt{K}$ is the 3×3 camera calibration matrix (cf. Hartley and Zisserman [1]):

$$\begin{bmatrix} f & 0 & p_x \\ 0 & f & p_y \\ 0 & 0 & 1 \end{bmatrix},$$

where $(p_x, p_y)^\top$ is the camera's principal point in pixel units. It follows that $\mathbf{P}$ is given in closed form by

$$Z \cdot \mathtt{K}^{-1} \begin{pmatrix} i \\ j \\ 1 \end{pmatrix} = Z \begin{bmatrix} 1/f & 0 & -p_x/f \\ 0 & 1/f & -p_y/f \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} i \\ j \\ 1 \end{pmatrix} = Z \begin{pmatrix} \frac{i-p_x}{f} \\ \frac{j-p_y}{f} \\ 1 \end{pmatrix}.$$

Horizontal Neighbors. We first consider the Euclidean distance δ_Z between two points both situated at depth Z and projecting to the horizontal pixel neighbors $({\color{cyan}i}, j)^\top, ({\color{red}i+1}, j)^\top$:

$$\begin{aligned} \delta_Z &= \left\| Z \cdot \mathtt{K}^{-1}({\color{red}i+1}, j, 1)^\top - Z \cdot \mathtt{K}^{-1}({\color{cyan}i}, j, 1)^\top \right\|_2 \\ &= \left\| Z \left(\mathtt{K}^{-1}({\color{red}i+1}, j, 1)^\top - \mathtt{K}^{-1}({\color{cyan}i}, j, 1)^\top \right) \right\|_2 \\ &= \left\| Z \begin{pmatrix} \frac{{\color{red}i+1}-p_x}{f} - \frac{{\color{cyan}i}-p_x}{f} \\ \frac{j-p_y}{f} - \frac{j-p_y}{f} \\ 1 - 1 \end{pmatrix} \right\|_2 = \left\| \begin{pmatrix} \frac{Z}{f} \\ 0 \\ 0 \end{pmatrix} \right\|_2 = \sqrt{\left(\frac{Z}{f} \right)^2} = Z/f. \end{aligned}$$

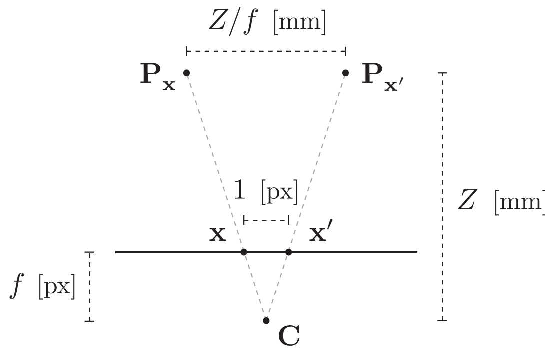

Figure 1: The Euclidean distance Z/f in world space between two points $\mathbf{P}_\mathbf{x}, \mathbf{P}_{\mathbf{x}'}$ both situated at depth Z—with respect to the camera center $\mathbf{C}$—and projecting to neighboring pixels $\mathbf{x}, \mathbf{x}'$.

Vertical Neighbors. Finally, we consider the analogous Euclidean distance δ'_Z for the vertical pixel neighbors $(i, j)^\top, (i, j+1)^\top$:

$$\begin{aligned}\delta'_Z &= \left\| Z \cdot \mathtt{K}^{-1}(i, j+1, 1)^\top - Z \cdot \mathtt{K}^{-1}(i, j, 1)^\top \right\|_2 \\ &= \left\| Z \left(\mathtt{K}^{-1}(i, j+1, 1)^\top - \mathtt{K}^{-1}(i, j, 1)^\top \right) \right\|_2 \\ &= \left\| Z \begin{pmatrix} \frac{i-p_x}{f} - \frac{i-p_x}{f} \\ \frac{j+1-p_y}{f} - \frac{j-p_y}{f} \\ 1-1 \end{pmatrix} \right\|_2 = \left\| \begin{pmatrix} 0 \\ \frac{Z}{f} \\ 0 \end{pmatrix} \right\|_2 = \sqrt{\left(\frac{Z}{f}\right)^2} = Z/f.\end{aligned}$$

References

[1] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2004.