3D-aware Image Editing for Out of Bounds Photography

Amit Shesh* Antonio Criminisi† Carsten Rother‡ Gavin Smyth§

Illinois State University Microsoft Research, Cambridge United Kingdom

Note this additional material is not important to understand the content of the paper

Appendix A: Calculating the focal length assuming a rectangular frame

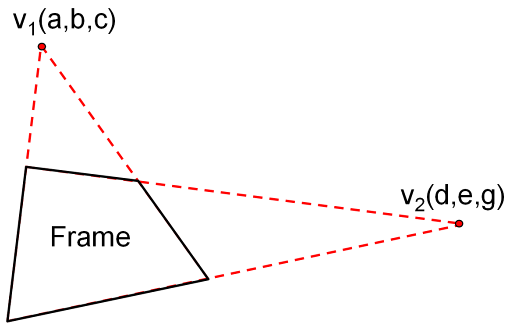

Figure 1: **A rectangular frame projected into the 2D image.**

In this section we describe how to calculate the focal length f of the camera assuming the frame to be rectangular. Since we set all other intrinsic camera parameters to default values, the full camera matrix K is then uniquely defined. The following is a detailed explanation of [HZ04] (Example 8.28 page 228).

Fig. 1 illustrates the scenario in the 2D image. Let $v_1 = \begin{pmatrix} a & b & c \end{pmatrix}^\top$ and $v_2 = \begin{pmatrix} d & e & g \end{pmatrix}^\top$ be the two vanishing points (in homogeneous coordinates) of the frame. We assume the intrinsic matrix K to be:

$$K = \begin{pmatrix} f & 0 & p_x \\ 0 & f & p_y \\ 0 & 0 & 1 \end{pmatrix}$$

where p_x and p_y are the coordinates of the principal point fixed at the center of the image, leaving f as the only unknown. Now, if we assume that the frame is rectangular then we obtain the linear constraint (see details in [HZ04])

$$v_1^T \omega v_2 = 0$$

where $\omega = \left(K^{-\top} K^{-1}\right) = \left(K K^\top\right)^{-1}$ is known as the absolute conic, an essential component for camera calibration. Therefore the linear constraint can be written explicitly as

$$\begin{pmatrix} a & b & c \end{pmatrix} \begin{pmatrix} \frac{1}{f^2} & 0 & \frac{-p_x}{f^2} \\ 0 & \frac{1}{f^2} & \frac{-p_y}{f^2} \\ \frac{-p_x}{f^2} & \frac{-p_y}{f^2} & \frac{p_x^2}{f^2} + \frac{p_y^2}{f^2} + 1 \end{pmatrix} \begin{pmatrix} d \\ e \\ g \end{pmatrix} = 0$$

and hence the focal f is given as

$$f = \sqrt{-\frac{a\left(d - g p_x\right) + b\left(e - g p_y\right) + c\left(g\left(p_x^2 + p_y^2\right) - d p_x - e p_y\right)}{cg}}.$$

In order to restrict ourselves to realistic cameras, we clamp the focal length to the range $[100, 3000]$. If the value of the focal length is a complex number we choose a standard value of $f = 750$.

*e-mail: ashesh@ilstu.edu
†e-mail:antcrim@microsoft.com
‡e-mail:carrot@microsoft.com
§e-mail:Gavin.Smyth@microsoft.com

Appendix B: Closed-form solution for frame placement

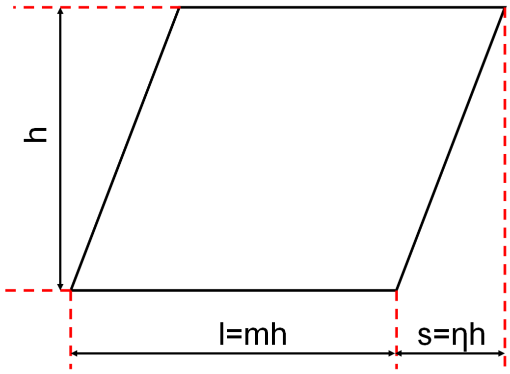

Figure 2: **Geometry of the frame in the X-Y plane.**

The goal of this subsection is to derive unique 3D coordinates for the four frame corners given in 2D. Let the frame lie in the X-Y plane in the world coordinate system as shown in fig. 2. For simplicity, let us express the geometry in terms of an aspect ratio m and a shear η. Thus, the values in the paper are given by $l = mh$ and $s = \eta h$. The four 3D vertices (in homogeneous coordinates) of the frame within this plane are given by

$$F_1 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, F_2 = \begin{pmatrix} mh \\ 0 \\ 1 \end{pmatrix}, F_3 = \begin{pmatrix} (m+\eta)h \\ h \\ 1 \end{pmatrix}, F_4 = \begin{pmatrix} \eta h \\ h \\ 1 \end{pmatrix} \tag{1}$$

Let K be the intrinsic camera matrix, which we assume to be known (from Appendix A). Let (p_{i1}, p_{i2}) be the rectified screen coordinates of the i^{th} vertex of the frame (i.e. for every 2D point $p = (x, y, 1)^\top$, the rectified point is given as $K^{-1}p$). Let $T_{camera} = [r_1|r_2|r_3|t]$ be the world-to-camera transformation matrix. It is a 3×4 matrix where $R = [r_1|r_2|r_3]$ is the camera's rotation. The rotation vectors have unit length ($|r_1|^2 = |r_2|^2 = |r_3|^2 = 1$) and are orthogonal to each other ($r_1 \perp r_2, r_2 \perp r_3, r_1 \perp r_3$), hence R has only 3 DOFs. Then the 2D points p_i and the 3D points F_i are related by

$$p_i \sim [r_1|r_2|t]F_i \tag{2}$$

where "$\sim$" means equality up to scale. The 3×3 matrix $[r_1|r_2|t]$ is a homography, which we write as H. Explicitly, the homography is

$$H = \begin{pmatrix} H_{11} & H_{12} & H_{13} \\ H_{21} & H_{22} & H_{23} \\ H_{31} & H_{32} & H_{33} \end{pmatrix} = [r_1|r_2|t]. \tag{3}$$

From eqn. 1-3 we obtain 8 linear independent equations on the unknown elements of H and the unknown frame parameters h, m, η:

$$\begin{aligned}
H_{33}p_{11} &= H_{13} \\
H_{33}p_{21} &= H_{23} \\
(mhH_{31} + H_{33})\,p_{12} &= mhH_{11} + H_{13} \\
(mhH_{31} + H_{33})\,p_{22} &= mhH_{21} + H_{23} \\
(h\,(m+\eta)\,H_{31} + hH_{32} + H_{33})\,p_{13} &= h\,(m+\eta)\,H_{11} + hH_{12} + H_{13} \\
(h\,(m+\eta)\,H_{31} + hH_{32} + H_{33})\,p_{23} &= h\,(m+\eta)\,H_{21} + hH_{22} + H_{23} \\
(h\eta H_{31} + hH_{32} + H_{33})\,p_{14} &= h\eta H_{11} + hH_{12} + H_{13} \\
(h\eta H_{31} + hH_{32} + H_{33})\,p_{24} &= h\eta H_{21} + hH_{22} + H_{23}
\end{aligned}$$

These constraints can be used to express H using the unknown parameters h, m, η, and the known screen coordinates p as:

$$H = \begin{pmatrix} H_{11} & H_{12} & H_{13} \\ H_{21} & H_{22} & H_{23} \\ H_{31} & H_{32} & H_{33} \end{pmatrix} = \begin{pmatrix} \frac{ek_1}{hmd} & \frac{emk_2+e\eta k_3}{-hmd} & ek_4 \\ \frac{ek_5}{hmd} & \frac{emk_6+e\eta k_7}{-hmd} & ek_8 \\ \frac{ek_9}{hmd} & \frac{emk_{10}+e\eta k_{11}}{-hmd} & e \end{pmatrix} \tag{4}$$

$$\begin{aligned}
k_1 &= p_{12}p_{13}p_{21} - p_{12}p_{14}p_{21} - p_{11}p_{13}p_{22} + p_{11}p_{14}p_{22} - p_{11}p_{14}p_{23} + p_{12}p_{14}p_{23} + p_{11}p_{13}p_{24} - p_{12}p_{13}p_{24} \\
k_2 &= -(p_{12}p_{14}p_{21} - p_{13}p_{14}p_{21} - p_{11}p_{13}p_{22} + p_{13}p_{14}p_{22} + p_{11}p_{12}p_{23} - p_{12}p_{14}p_{23} - p_{11}p_{12}p_{24} + p_{11}p_{13}p_{24}) \\
k_3 &= -(-p_{12}p_{13}p_{21} + p_{12}p_{14}p_{21} + p_{11}p_{13}p_{22} - p_{11}p_{14}p_{22} + p_{11}p_{14}p_{23} - p_{12}p_{14}p_{23} - p_{11}p_{13}p_{24} + p_{12}p_{13}p_{24}) \\
k_4 &= p_{11} \\
k_5 &= -(-p_{12}p_{21}p_{23} + p_{14}p_{21}p_{23} + p_{11}p_{22}p_{23} - p_{14}p_{22}p_{23} + p_{12}p_{21}p_{24} - p_{13}p_{21}p_{24} - p_{11}p_{22}p_{24} + p_{13}p_{22}p_{24}) \\
k_6 &= -(-p_{13}p_{21}p_{22} + p_{14}p_{21}p_{22} + p_{12}p_{21}p_{23} - p_{14}p_{21}p_{23} - p_{11}p_{22}p_{24} + p_{13}p_{22}p_{24} + p_{11}p_{23}p_{24} - p_{12}p_{23}p_{24}) \\
k_7 &= -(-p_{12}p_{21}p_{23} + p_{14}p_{21}p_{23} + p_{11}p_{22}p_{23} - p_{14}p_{22}p_{23} + p_{12}p_{21}p_{24} - p_{13}p_{21}p_{24} - p_{11}p_{22}p_{24} + p_{13}p_{22}p_{24}) \\
k_8 &= p_{21} \\
k_9 &= p_{13}p_{21} - p_{14}p_{21} - p_{13}p_{22} + p_{14}p_{22} - p_{11}p_{23} + p_{12}p_{23} + p_{11}p_{24} - p_{12}p_{24} \\
k_{10} &= -p_{12}p_{21} + p_{13}p_{21} + p_{11}p_{22} - p_{14}p_{22} - p_{11}p_{23} + p_{14}p_{23} + p_{12}p_{24} - p_{13}p_{24} \\
k_{11} &= p_{13}p_{21} - p_{14}p_{21} - p_{13}p_{22} + p_{14}p_{22} - p_{11}p_{23} + p_{12}p_{23} + p_{11}p_{24} - p_{12}p_{24} \\
d &= p_{13}p_{22} - p_{14}p_{22} - p_{12}p_{23} + p_{14}p_{23} + p_{12}p_{24} - p_{13}p_{24}
\end{aligned}$$

As eqn. 2 is true up to a scale of H, we fix this scale arbitrarily by assigning $e = 1$.
Imposing the constraint $r_1 \perp r_2$, we get:

$$H_{11}H_{12} + H_{21}H_{22} + H_{31}H_{32} = 0$$

This defines the relationship between η and m as

$$\eta = \beta m \ , \ \ \beta = -\frac{k_1k_2 + k_5k_6 + k_9k_{10}}{k_1k_3 + k_5k_7 + k_9k_{11}}. \tag{5}$$

Next we impose the constraint $|r_1|^2 = |r_2|^2$ which gives

$$H_{11}^2 + H_{21}^2 + H_{31}^2 = H_{12}^2 + H_{22}^2 + H_{32}^2.$$

This uniquely specifies m as

$$m = \sqrt{\frac{{k_1}^2 + {k_5}^2 + {k_9}^2}{(k_2 + \beta k_3)^2 + (k_6 + \beta k_7)^2 + (k_{10} + \beta k_{11})^2}} \tag{6}$$

and hence η is defined using eqn. 5. Note, eqn. 5 and eqn. 6 always provide a real and positive value for m and a real value for η.

To get the value for h, we apply the constraint that $|r_1|^2$=1:

$$\left(\frac{k_1}{hmd}\right)^2 + \left(\frac{k_5}{hmd}\right)^2 + \left(\frac{k_9}{hmd}\right)^2 = 1$$

which gives h as

$$h = \sqrt{\frac{{k_1}^2 + {k_5}^2 + {k_9}^2}{m^2d^2}} \tag{7}$$

Finally, we obtain our full camera matrix T_{camera} from eqn. 4 (since $H = [r_1|r_2|t]$) and $r_3 = r_1 \times r_2$.

References

[HZ04] Hartley R. I., Zisserman A.: *Multiple View Geometry in Computer Vision*, second ed. Cambridge University Press, ISBN: 0521540518, 2004.