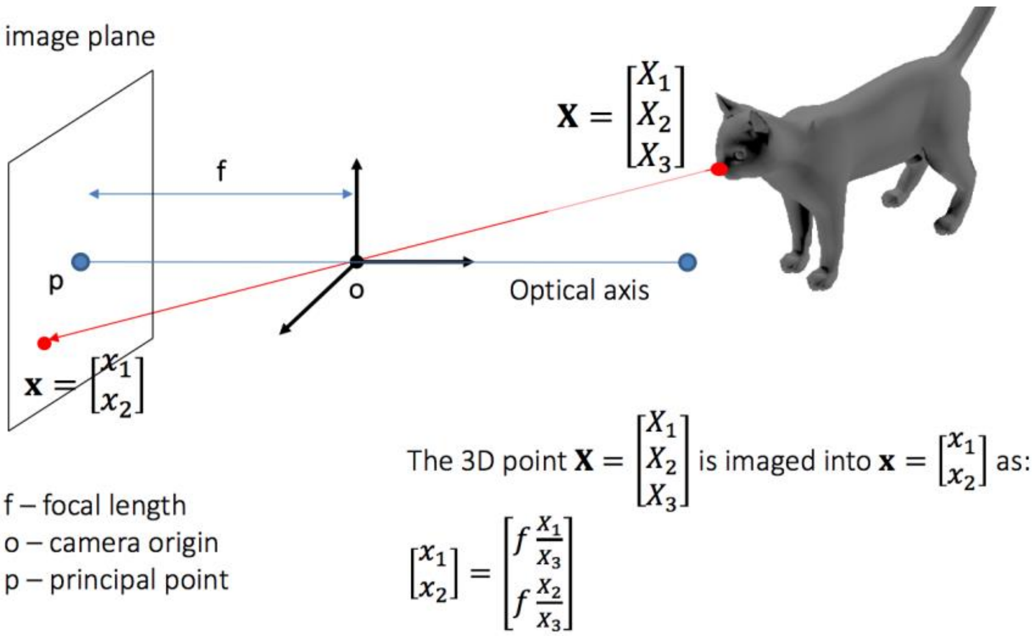
Exercise 1. Pinhole camera.

The perspective projection equations for a pinhole camera are

$$x_p = f\frac{x_c}{z_c}, \qquad y_p = f\frac{y_c}{z_c}, \tag{1}$$

where $\mathbf{x}_p = [x_p, y_p]^\top$ are the projected coordinates on the image plane, $\mathbf{x}_c = [x_c, y_c, z_c]^\top$ is the imaged point in the camera coordinate frame and f is the focal length. Give a geometric justification for the perspective projection equations.
(Hint: Use similar triangles and remember that the image plane is located at a distance f from the projection center and is perpendicular to the optical axis, i.e. the z-axis of the camera coordinate frame.)



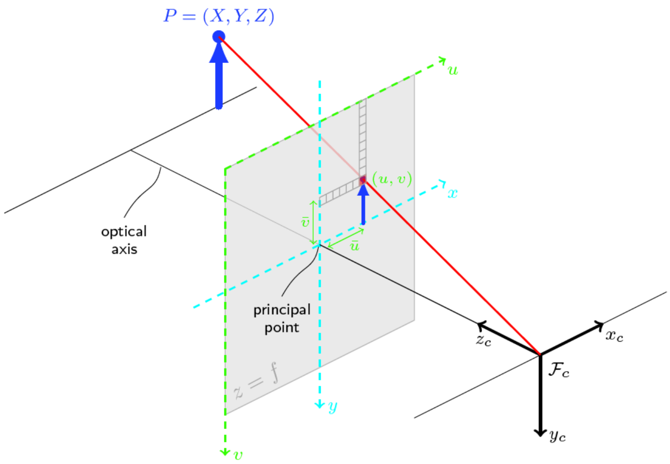
Let’s denote O as the camera origin, p as the principle point on the image plane, x is the projected coordinated, X as the original point in 3D world and P as the original principle point lying on the same plane with X and this plane is parallel to the image plane.

We see that the triangles Opx and the triangle OPX are similar triangles. Therefore, the ratio between the edges should be equal, or $\frac{OP}{Op} = \frac{x_c}{x_p} = \frac{y_c}{y_p} \Rightarrow \frac{x_c}{x_p} = \frac{y_c}{y_p} = \frac{z_c}{f}$. Finally, the 3D coordinate is projected onto the plane as $\left(f\frac{x_c}{z_c}, f\frac{y_c}{z_c}, f \right) \Rightarrow (x_p, y_p, z_p)$, where the third coordinate is omitted as the projected plane is 2D Euclidean frame. The projected coordinate point is thus derived as:

$$(x_p, y_p) = \left(f\frac{x_c}{z_c}, f\frac{y_c}{z_c} \right)$$

Exercise 2. Pixel coordinate frame.
The image coordinates x_p and y_p given by the perspective projection equations (1) above are not in pixel units. The x_p and y_p coordinates have the same unit as distance f (typically millimeters) and the origin of the coordinate frame is the principal point (the point where the optical axis pierces the image plane). Now, give a formula which transforms the point $\mathbf{x}_p$ to its pixel coordinates $\mathbf{p} = [u, v]^\top$ when the number of pixels per unit distance in u and v directions are m_u and m_v, respectively, the pixel coordinates of the principal point are (u_0, v_0) and
a) u and v axis are parallel to x and y axis, respectively.
b) u axis is parallel to x axis and the angle between u and v axis is θ.

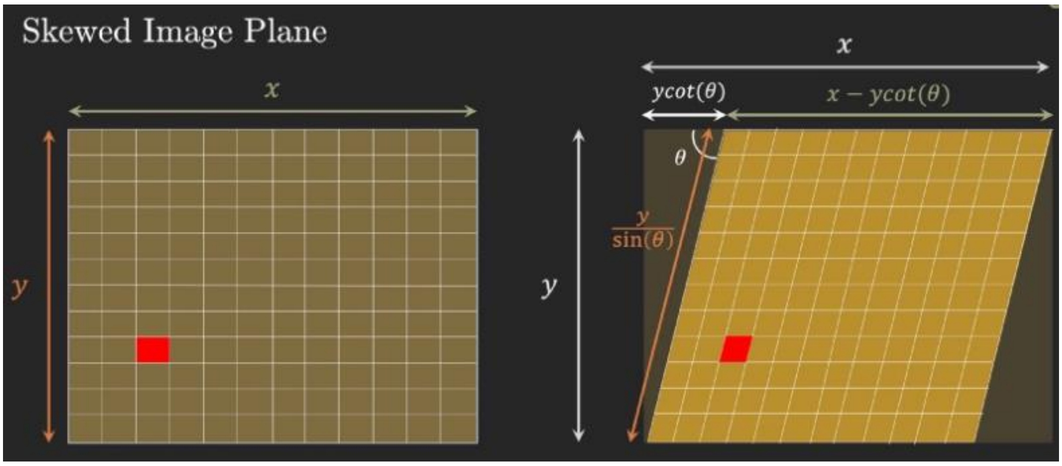
a) The case when u and v axis are parallel to x and y axis, respectively



We can apply the same pinhole camera equation from Exercise 1, except now that the focal lengths should be expressed in terms of pixels, not in the unit length. We know that there are m_u and m_v number of pixels in each unit in their respective directions, so the focal lengths in pixels are $f_x = m_u f$ and $f_y = m_v f$. Since the size of the pixels are not necessarily rectangular, the focal lengths in both direction will be different with the formula given above. Additionally, the principal point is no longer (0, 0) like in Exercise 1, but expressed as (u_0, v_0), so all pixel coordinates should be shifted by this principal point. Plugging into the equations, we have the pixel coordinate calculated from the camera coordinate and the focal length as:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} f_x \dfrac{x_c}{z_c} + u_0 \\ f_y \dfrac{y_c}{z_c} + v_0 \end{bmatrix} = \begin{bmatrix} m_u f \dfrac{x_c}{z_c} + u_0 \\ m_v f \dfrac{y_c}{z_c} + v_0 \end{bmatrix} = \begin{bmatrix} m_u x_p + u_0 \\ m_v y_p + v_0 \end{bmatrix} \text{(answer)}$$

b) The case when u axis is parallel to the x-axis and the angle between u and v axis is θ



Since the image plane (pixel coordinate) is skewed with angle θ, the skewed projected coordinate is:

$$\begin{bmatrix} x_{p_{skewed}} \\ y_{p_{skewed}} \end{bmatrix} = \begin{bmatrix} 1 & -\cot\theta \\ 0 & \dfrac{1}{\sin\theta} \end{bmatrix} \begin{bmatrix} x_p \\ y_p \end{bmatrix} \Rightarrow \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} m_u\left(x_p - y_p\cot\theta\right) + u_0 \\ m_v\left(\dfrac{\sin\theta}{y_p}\right) + v_0 \end{bmatrix} \text{(answer)}$$

Exercise 3. Intrinsic camera calibration matrix.
Use homogeneous coordinates to represent case (2.a) above with a matrix $\mathbf{K}_{3\times3}$, also known as the camera's intrinsic calibration matrix, so that $\tilde{\mathbf{p}} = \mathbf{K}\mathbf{x}_c$. Where $\tilde{\mathbf{p}}$ is $\mathbf{p}$ in homogeneous coordinates.

From exercise (2), we have defined the equations for the UV coordinates. Since $\tilde{p}$ is homogeneous coordinates, it can be scaled by any factor. Since the z coordinate of x_c is z_c, $\tilde{p}$ will be scaled by z_c:

$$\mathrm{x}_c = \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} \Rightarrow \tilde{p} = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} z_c u \\ z_c v \\ z_c \end{bmatrix} = \begin{bmatrix} z_c\left(f_x \dfrac{x_c}{z_c} + u_0\right) \\ z_c\left(f_y \dfrac{y_c}{z_c} + v_0\right) \\ z_c \end{bmatrix} = \begin{bmatrix} f_x x_c + z_c u_0 \\ f_y y_c + z_c v_0 \\ z_c \end{bmatrix}$$

We need to find $K_{3\times3}$ such that $\tilde{p} = K\mathrm{x}_c$. In the first row of $\tilde{p}$, there is x_c and z_c but no y_c. In the second row, there is y_c and z_c but no x_c. In the third row, there is only z_c. The coefficients are known and therefore, we can construct the intrinsic camera calibration matrix as:

$$K = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$ (answer). Plugging in, we have: $\tilde{p} = K\mathrm{x}_c$ or $$\begin{bmatrix} z_c u \\ z_c v \\ z_c \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix}$$

Exercise 4. Camera projection matrix.
Imaged points are often expressed in an arbitrary frame of reference called the world coordinate frame. The mapping from the world frame to the camera coordinate frame is a rigid transformation consisting of a 3D rotation $\mathbf{R}$ and translation $\mathbf{t}$:

$$\mathbf{x}_c = \mathbf{R}\mathbf{x}_w + \mathbf{t}.$$

Use homogeneous coordinates and the result of the exercise 3 above, to write down the 3×4 camera projection matrix $\mathbf{P}$ that projects a point form world coordinates $\mathbf{x}_w$ to pixel coordinates. That is, represent $\mathbf{P}$ as a function of the internal camera parameters $\mathbf{K}$ and the external camera parameters $\mathbf{R}, \mathbf{t}$.

The extrinsic parameter calibration matrix is the concatenation of the rotation and translation matrices

$$E = [R \mid t] = \left[\begin{array}{ccc|c} \cos\theta & -\sin\theta & 0 & t_x \\ \sin\theta & \cos\theta & 0 & t_y \\ 0 & 0 & 1 & 1 \end{array}\right]$$. This translates the world coordinate to the camera coordinate

From exercise (3), the intrinsic paramter calibration matrix is $K = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix}$

The 3x4 camera projection matrix P that projects the world from world coordinates x_w to pixel coordinates can be formulated by multiply K and E together:

$$P = KE = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta & 0 & t_x \\ \sin\theta & \cos\theta & 0 & t_y \\ 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} f_x\cos\theta & -f_x\sin\theta & u_0 & f_x t_x \\ f_y\sin\theta & f_y\cos\theta & v_0 & f_y t_y \\ 0 & 0 & 1 & 1 \end{bmatrix} \text{ (answer)}$$

In other words, the projection from world coordinates to the pixel coordinates is:

$$\tilde{p} = P\mathrm{x}_w \Rightarrow z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x\cos\theta & -f_x\sin\theta & u_0 & f_x t_x \\ f_y\sin\theta & f_y\cos\theta & v_0 & f_y t_y \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix}$$

Exercise 5. Rotation matrix.
A rigid coordinate transformation can be represented with a rotation matrix $\mathbf{R}$ and a translation vector $\mathbf{t}$, which transform a point $\mathbf{x}$ to $\mathbf{x}' = \mathbf{R}\mathbf{x} + \mathbf{t}$. Now, let the 3×3 matrix $\mathbf{R}$ be a 3-D rotation matrix, which rotates a vector $\mathbf{x}$ by the angle θ about the axis $\mathbf{u}$ (a unit vector). According to the Rodrigues formula it holds that

$$\mathbf{R}\mathbf{x} = \cos\theta\,\mathbf{x} + \sin\theta\,\mathbf{u}\times\mathbf{x} + (1-\cos\theta)(\mathbf{u}\cdot\mathbf{x})\mathbf{u}.$$

a) Give a geometric justification (i.e. derivation) for the Rodrigues formula.

b) Derive the expressions for the elements of $\mathbf{R}$ as a function of θ and the elements of $\mathbf{u}$. (Hint: Write down the cross product $\mathbf{u}\times\mathbf{x}$ and the scalar product $\mathbf{u}\cdot\mathbf{x}$ in terms of the elements of vectors $\mathbf{u}$ and $\mathbf{x}$. Use the notations $\mathbf{u} = (u_1, u_2, u_3)^\top$ and $\mathbf{x} = (x_1, x_2, x_3)^\top$. You may also consult literature or public sources like Wikipedia.)

a) The geometric derivation for the Rodrigues formula

The vector x can be decomposed into components parallel and perpendicular to the axis u as:

⇨ $x = x_{\parallel} + x_{\perp}$

The component parallel to u is the vector projection of x on u, which can be derived as:

⇨ $x_{\parallel} = (x\cdot u)u$

The component perpendicular to u is the vector rejection of x on u, which can be derived as:

⇨ $x_{\parallel} = x - x_{\parallel} = x - (u\cdot v)u = -u\times(u\times x)$

The component parallel to the axis does not change magnitude nor direction under rotation

⇨ $x_{\parallel rot} = x_{\parallel}$

Only the perpendicular component changes its direction. Its magnitude, however, stays the same under rotation.

⇨ $|x_{\perp rot}| = x_{\perp}$ and $x_{\perp rot} = \cos(\theta)x_{\perp} + \sin(\theta)u\times x_{\perp}$ (1)

Because u and $x_{\parallel}$ are parallel, their cross product is zero, which leads to:

⇨ $u\times x_{\perp} = u\times(x - x_{\parallel}) = u\times x - u\times x_{\parallel} = u\times x$

From (1), it follows as: $x_{\perp rot} = \cos(\theta)x_{\perp} + \sin(\theta)u\times x$. Finally, the fully rotated vector is:

$x_{rot} = x_{\parallel rot} + x_{\perp rot}$ (2). Substitute the identities $x_{\parallel rot}$ and $x_{\perp rot}$ above into (2), we have:

$$\begin{aligned}
x_{rot} &= x_{\parallel} + \cos(\theta)x_{\perp} + \sin(\theta)u\times x \\
&= x_{\parallel} + \cos(\theta)(x - x_{\parallel}) + \sin(\theta)u\times x \\
&= \cos(\theta)x + (1-\cos(\theta))x_{\parallel} + \sin(\theta)u\times x \\
&= \cos(\theta)x + (1-\cos(\theta))(u\cdot x)u + \sin(\theta)u\times x \text{ (Proven)}
\end{aligned}$$

b) Derive the expressions for the element of R as a function of θ and u.

We have $Rx = x\cos\theta + (1-\cos\theta)(u\cdot x)u + \sin\theta(u\times x)$ from part (a)

$$\Rightarrow \quad R\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \cos\theta \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + (1-\cos\theta)\begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}\begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}\begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} + \sin\theta\left(\begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}\times\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right)$$

$$\Rightarrow \quad R\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1\cos\theta + (1-\cos\theta)\left(x_1u_1^2 + x_2u_1u_2 + x_3u_1u_3\right) + \sin\theta\left(x_3u_2 - x_2u_3\right) \\ x_2\cos\theta + (1-\cos\theta)\left(x_1u_1u_2 + x_2u_2^2 + +x_3u_2u_3\right) + \sin\theta\left(x_1u_3 - x_3u_1\right) \\ x_3\cos\theta + (1-\cos\theta)\left(x_1u_1u_3 + x_2u_2u_3 + x_3u_3^2\right) + \sin\theta\left(x_2u_1 - x_1u_2\right) \end{bmatrix}$$

Extracting out the coefficients for each row, R can finally be derived as:

$$R = \begin{bmatrix} \cos\theta + u_1^2(1-\cos\theta) & u_1u_2(1-\cos\theta) - u_3\sin\theta & u_1u_3(1-\cos\theta) + u_2\sin\theta \\ u_2u_1(1-\cos\theta) + u_3\sin\theta & \cos\theta + u_2^2(1-\cos\theta) & u_2u_3(1-\cos\theta) - u_1\sin\theta \\ u_3u_1(1-\cos\theta) - u_2\sin\theta & u_3u_2(1-\cos\theta) + u_1\sin\theta & \cos\theta + u_3^2(1-\cos\theta) \end{bmatrix} \text{(answer)}$$