

The objective of motion of mobile camera (mounted on a robotic arm) is to focus on detected ear from images acquired with a fixed camera. We assume that intrinsic matrices and camera planes are the same for the computation. In a first step, mobile camera moves parallel to the camera plane in order to position the selected object (i.e. detected ear) in the center of the image. Then, we translate the camera in the plant direction to acquire an image of better resolution.

Let (u, v) be the pixel position in the image and (X, Y, Z) a 3D point in the world coordinate frame of the scene. The projection (u, v) of the point in the image is defined as:

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

with (f_x, f_y) the focal length, (c_x, c_y) the center of the image in pixel, and R the transformation matrix between the camera frame and the global coordinate frame.

If we define the camera frame as the global coordinate frame, then the projection of a pixel (u_1, v_1) can be expressed as :

$$s \begin{bmatrix} u_1 \\ v_1 \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \Rightarrow \begin{bmatrix} s u_1 \\ s v_1 \\ s \end{bmatrix} = \begin{bmatrix} X f_x + Z c_x \\ Y f_y + Z c_y \\ Z \end{bmatrix}$$

X and Y object coordinates can be expressed as :

$$\begin{cases} X = \frac{Z(u_1 - c_x)}{f_x} \\ Y = \frac{Z(v_1 - c_y)}{f_y} \end{cases} \quad (1)$$

In the previous frame, we translate the camera frame to position the projection of (X, Y, Z) in the center of the image (u_2, v_2) :

$$s \begin{bmatrix} u_2 \\ v_2 \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & t_y \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad \text{with} \quad \begin{cases} u_2 = \frac{\text{width}}{2} \\ v_2 = \frac{\text{height}}{2} \end{cases}$$

X and Y object coordinates can be expressed as :

$$\begin{cases} X = \frac{Z(u_2 - c_x)}{f_x} - t_x \\ Y = \frac{Z(v_2 - c_y)}{f_y} - t_y \end{cases} \quad (2)$$

Combination of (1) and (2) gives the expression of t_x and t_y :

$$\begin{cases} t_x = \frac{Z(u_2 - u_1)}{f_x} \\ t_y = \frac{Z(v_2 - v_1)}{f_y} \end{cases}$$

The extrinsic matrix is that it describes how the world is transformed relative to the camera, so the camera's target position is the opposite of calculated t_x and t_y .

The transformation matrix from fixed camera to mobile camera is a rotation matrix (fig. 1) :

$$\begin{bmatrix} X_a \\ Y_a \\ Z_a \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} + \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \end{bmatrix} \Rightarrow \begin{bmatrix} X_a \\ Y_a \\ Z_a \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} -t_x \\ -t_y \\ 0 \end{bmatrix} + \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \end{bmatrix}$$

The target position to pilot the mobile camera is finally :

$$\begin{cases} X_a = [\text{depth, depending on desired resolution}] + X_0 \\ Y_a = \frac{Z(u_1 - u_2)}{f_x} + Y_0 \\ Z_a = \frac{Z(v_2 - v_1)}{f_y} + Z_0 \end{cases}$$

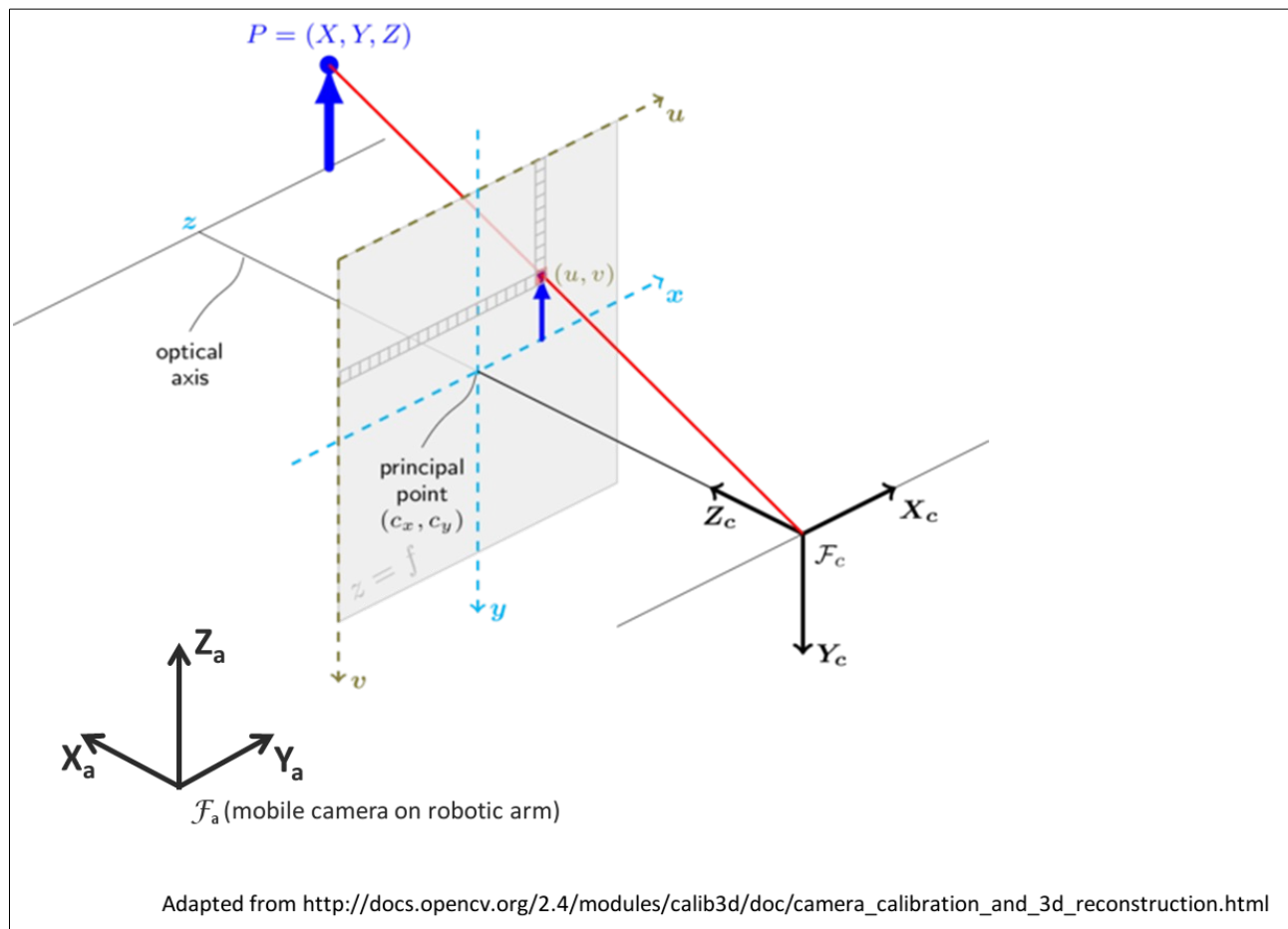


Fig.1: Image, fixed camera and mobile camera coordinates in imaging cabin