2. [8] **Rigid Body Transformation**
Find the rigid transformation from a scene coordinate frame to a camera coordinate frame by specifying the 4 × 4 matrix relating a scene point given by the homogeneous representation of its world coordinates, (X, Y, Z, 1), to its homogeneous representation in camera coordinates, (x, y, z, 1), for the arrangement shown in the figure in Problem 1.

Rigid body transformations can be described by

$$\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & T_x \\ r_{21} & r_{22} & r_{23} & T_y \\ r_{31} & r_{32} & r_{33} & T_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

Here, the relation between the scene and camera coordinate frames is a translation by (0, 75, 500), a rotation about the x-axis by +90 degrees, and a rotation about the z-axis by +90 degrees. So, $T_x = 0$, $T_y = 75$, and $T_z = 500$. The 3 × 3 R matrix for rotation by 90 degrees about the x-axis is

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}$$

Similarly, the 3 × 3 R matrix for rotation by 90 degrees about the z-axis is

$$\begin{bmatrix} \cos\theta & -sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Composing these two rotation matrices, we get

$$R = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & -1 \\ 1 & 0 & 0 \end{bmatrix}$$

So, the final 4 × 4 matrix is

$$\begin{bmatrix} 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 75 \\ 1 & 0 & 0 & 500 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

3. [9] **Projective Cameras**

The relationship between a 3D world point (X, Y, Z) and its 2D image pixel coordinates (u, v) can be written as:

$$\begin{bmatrix} su \\ sv \\ s \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

(a) [3] Under what assumptions or constraints is this relationship valid?

The above general 3 × 4 matrix P defines a projective camera, which is a generalization of a perspective camera, which models the true relationship between a world point and an image point. That is, a perspective camera constrains P to be defined as P = K[R|T], where K is the 3 × 3 camera calibration matrix, R is the 3 × 3 rotation matrix, and T is the 3 × 1 translation matrix.

(b) [3] Why is this projective camera often preferred over the perspective camera model?

Because the projective camera is linear, without the nonlinear constraints on the elements of P due to perspective.

(c) [3] How should the relationship be modified to represent points at infinity in both the image plane and the world?

A scene point at infinity is represented in homogeneous coordinates as (X, Y, Z, 0) and an image point at infinity is represented as (u, v, 0). This means the last column of P can be eliminated, making P 3 × 3:

$$\begin{bmatrix} u \\ v \\ 0 \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$