Problem 2:
Sensor chosen: A RGB camera mounted on the robot end effector flange to obtain visual information of the environment.

Notations:
m: Model coordinate frame
c: Camera coordinate frame
e: Tool/end-effector coordinate frame
w: World coordinate frame in which robot base is located

A coordinate transformation T_{ab} denotes a 4x4 homogeneous transformation matrix which describes the position of frame 'b' origin and the orientation of axes relative to frame 'a'. Similarly the following holds true:

T_{cm}: Transformation from model frame to camera frame
T_{ec}: Transformation from camera frame to robot end effector frame
T_{we}: Transformation from end effector/tool frame to world frame(robot origin)
T_{wm}: Transformation from model frame to world/robot base frame

The metadata of the point of interest contains the coordinates in the model's coordinate frame, the goal here is to establish a mapping between the model coordinate frame and camera center/coordinate frame. Assuming these points are marked on the object as well, we can extract these features using feature detectors and compute 2D points (image points) of these features in the image. This boils down to establishing 3D-2D correspondences.

Solving for the projection matrix gives the pose of the model wrt to the camera's coordinate frame.

Consider the world frame is fixed to the robot's base frame.The vision system involves 3D-2D projection and mapping and it can be achieved using a pinhole camera model.

$$[u,v,1]^T = H_{cm}[x_m,y_m,z_m]^T \qquad (1)$$

Where:

- H_{cm} is the rigid body transformation from model coordinate frame to camera coordinate frame.
- x_m,y_m,z_m are the coordinates of points in the model coordinate frame and u,v are the projection of points onto the image plane.

The next step would be to compute the transformation between the camera and the end effector.

The following equation helps to achieve the pose transform between the model coordinate frame to robot base coordinate frame:

$$T\_wm = T\_cm \cdot T\_ec \cdot T\_we \qquad (2)$$

- The coordinate transformation **T_we** can be calculated using joint angles and kinematic equations.
- **T_cm** can also be computed using perspective projection since the model coordinates in 3D are known.
- The process of estimating the transformation between the camera frame and the end effector frame **T_ec** is known as hand eye calibration.

  For solving the hand eye calibration problem, we can use a regular calibration pattern like the AruCo marker as a target. The 3D-2D correspondence between the 3D points on the target and 2D images can be obtained using perspective projection for every pose of the camera. The transform from the base of the robot arm and the end effector can be obtained from robot kinematics. The calibration process consists of moving the arm through a series of poses. At each pose we can get a pair of transforms: Base to end effector transform and the transform b/w camera and target (AruCo). Then these poses can be put together in **AX = XB** equation to recover the camera to the end effector transform.

  The above equation is solved using open source implementations which give the transformation of camera frame wrt to the end effector frame.