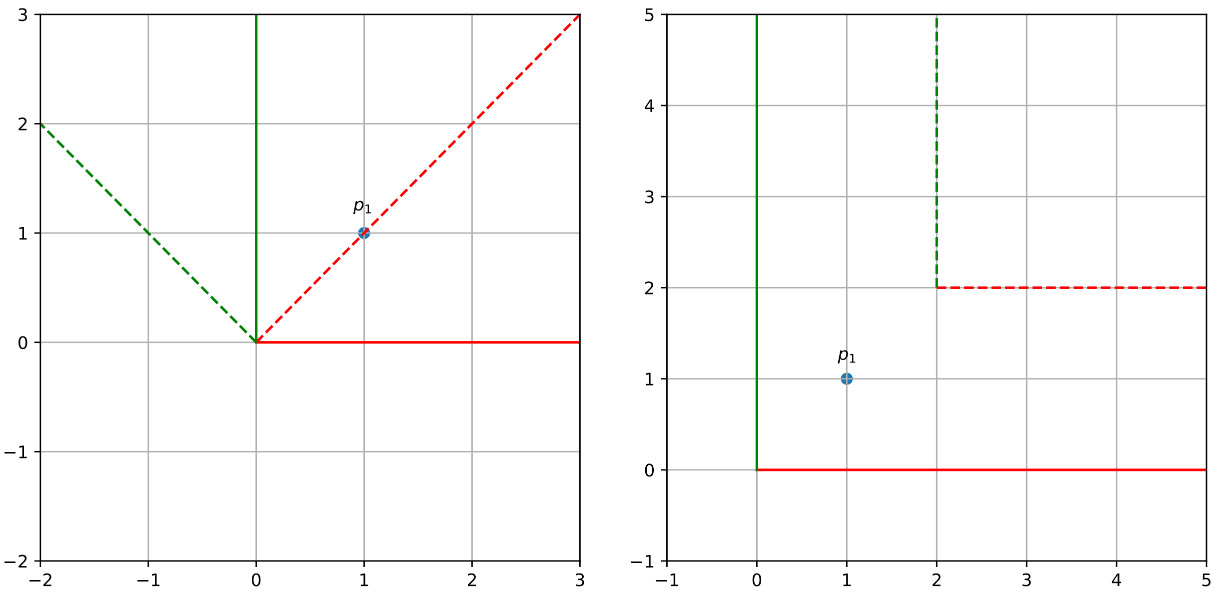
Technische Universität München
Fakultät für Informatik
Yehya Abouelnaga, Ivan Pavlov
and Adrian Haarbach

WS 19/20
Solution for Exercise Sheet 1
November 7, 2019

Exercises in Tracking & Detection

Task 1 Transformations in 2D Plane



(a) Rotated Coordinate System (b) Translated Coordinate System

Figure 1: The above figure shows three coordinate systems: the world coordinate system (solid line in both sub-figures), a rotated coordinate system (Left; counterclockwise 45 degrees), a translated coordinate system (Right). The world coordinate is shown by the non-dashed lines in both graphs. The red lines refer to the X-Axis. The green lines refer to the Y-Axis. Point p_1 is defined as $(1, 1)$ in the world coordinate system.

- Write the transformation from the world coordinate system to the local coordinate system in Figure 1a. Re-compute the point p_1 in the local coordinate system (i.e. the rotated one).

Answer

$$p_W = T_{L\to W} \cdot p_L = [\sqrt{2}, 0]^T$$

$$T_{L->W} = [\mathbf{i}' \ \mathbf{j}'] = \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}$$

where $\mathbf{i}', \mathbf{j}'$ are axes of the local coordinate system expressed in World coordinates. Since the local coordinate system has only been rotated this transformation matrix is rotation matrix and has properties of the the rotation matrix. Having that in mind we know that $R^T = R^{-1}$, $R^T R = RR^T$, so we can write the opposite change of the coordinate systems:

$$T_{W->L} = (T_{L->W})^{-1} = (T_{L->W})^T = \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}$$

so the point p_L in a local coordinate system can be written as:

$$p_L = T_{W->L} \cdot p_W = \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = [\sqrt{2}, 0]^T$$

- Write the transformation from the world coordinate system to the local coordinate system at Figure 1b. Re-compute the point p_1 in the local coordinate system (i.e. the translated one).

Answer For translation, we need to express the point p (either in local p_L or world p_W coordinates) using homogeneous coordinates:

$$p_W = T_{L\to W} \cdot p_L$$

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix} = T_{L\to W} \cdot \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}$$

$$T_{L->W} = \begin{bmatrix} R & t \end{bmatrix} = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 2 \end{bmatrix}$$

Tranformation matrix $T_{L->W}$ is composed of rotational part that is identity matrix, since the world and local coordinate system are aligned and rotation between them doesn't exists, and the translational part described by the last column represents coordinates of the origin of the local coordinate system in the world coordinates. The opposite transformation from world to local is inverse transformation.

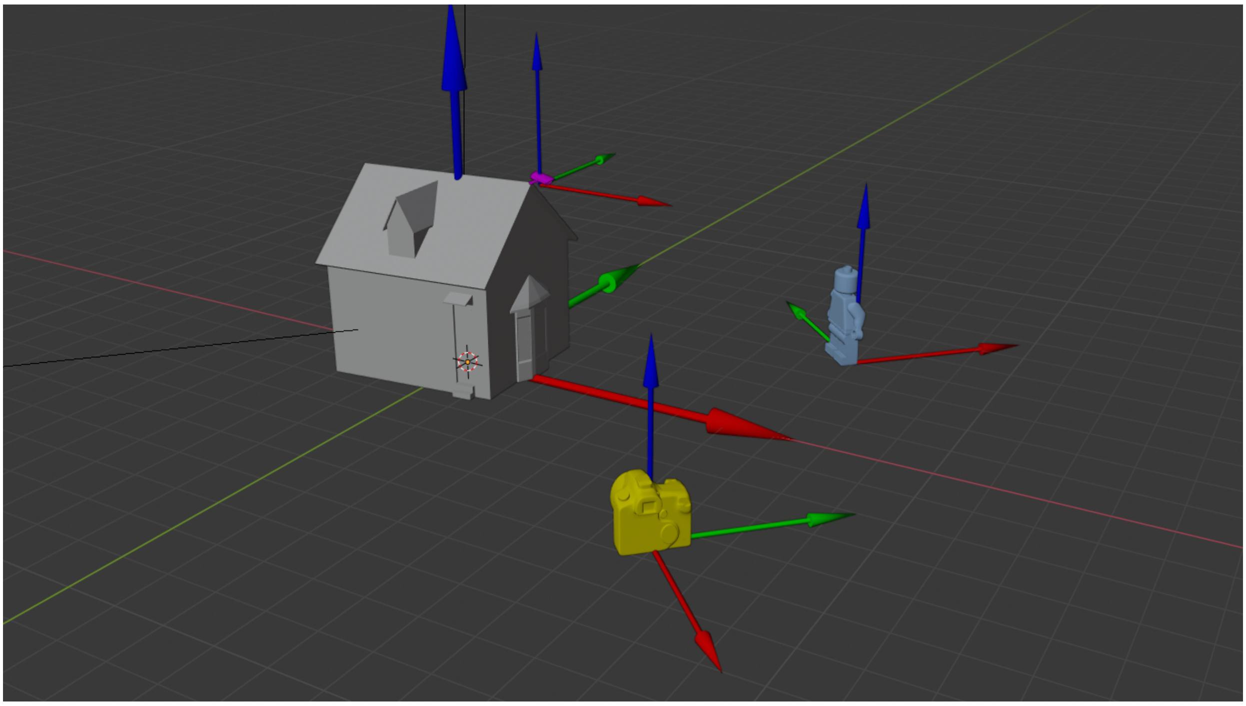
$$T_{W->L} = (T_{L->W})^{-1} = \begin{bmatrix} R^T & -R^T t \end{bmatrix} = \begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & -2 \end{bmatrix}$$

$$p_L = T_{W->L} \cdot p_W = \begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & -2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = [-1, -1]^T$$

Note that the above computation can be also obtained if the transformation matices are written with their homogenoious part (the third raw being $[0, 0, 1]$). Above we use compact notation of 2×3 transformation matrices that is eqvivalent to 3×4 transformation matrices in 3D.

Task 2 House Robber: Transformations in 3D

In the above figure, we have the following objects: House, CCTV Camera on top of the house, a robber (i.e. the robot) and a neighbor's camera (the orange camera). The scene assumes a right-handed coordinate system. The red and green arrows correspond to the X- and Y-Axes, respectively. The Z-axis (in blue) is orthogonal to the X- and Y-Axes. The robot is on the positive side of the X- and Y-Axes. The house center lies in the origin of the world coordinate system (i.e. look at the red circle).



- What is the transformation (i.e. pose) of the CCTV camera relative to the house? The house dimensions are 2m, 2m and 4m along the X-, Y- and Z-Axes. Assume that the X-Axis and the Y-Axis divide the house into symmetrical halves.

  Answer Given that X- and Y-Axis divide the house into symmetrical halves, we know that to reach the edge of the house on the X-Axis, we need to translate 1m in the positive direciton of the X-Axis. Since the CCTV camera lays on the Y-axis, there is no translation along Y. While there is no rotation applied to the CCTV camera, we use the Identity matrix as our rotation.

$$T_{CCTV \to World} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

  The 4×4 matrix $T_{CCTV \to World}$ describes the pose of the CCTV camera in the world coordinates. This pose could be used to transform a point from the CCTV local coordinate system into the world coordinate system.

- What is the transformation of the robot (i.e. its pose) relative to the house? The robot is translated 6m on the positive direction of the X-Axis and 5m on the positive direction of the Y-Axis in the world coordinates. The default object model (for the robot 3D model) is desigend to look at the negative X-Axis (when no rotation is applied). In order for it to face the house, we rotate it 45 degrees around the Z-Axis.

  Answer Rotation defined by angles of rotation γ, β, α around the X-,Y-, and Z-axes could be represented by the following rotation matrix:

$$\mathbf{R} = \underbrace{\begin{bmatrix} \cos(\alpha) & -\sin(\alpha) & 0 \\ \sin(\alpha) & \cos(\alpha) & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{Z\text{-Axis Rotation}} \underbrace{\begin{bmatrix} \cos(\beta) & 0 & \sin(\beta) \\ 0 & 1 & 0 \\ -\sin(\beta) & 0 & \cos(\beta) \end{bmatrix}}_{Y\text{-Axis Rotation}} \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\gamma) & -\sin(\gamma) \\ 0 & \sin(\gamma) & \cos(\gamma) \end{bmatrix}}_{X\text{-Axis Rotation}}$$

R rotates a coordinate system γ degrees around the X-Axis, then β degrees around the Y-Axis, then finally α degrees around Z-Axis. **Beware that order of multiplication matters here!**

In order to formulate the robot pose in the world coordinate, we use the following transformation matrix:

$$T_{Robot \to World} = \begin{bmatrix} cos(\frac{\pi}{4}) & -sin(\frac{\pi}{4}) & 0 & 6 \\ sin(\frac{\pi}{4}) & cos(\frac{\pi}{4}) & 0 & 5 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} & 0 & 6 \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 & 5 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

- What is the robot transformation (i.e. its pose) relative to the CCTV camera?

Answer In order to express the robot pose relative to the CCTV camera, we need to find a sequence of transformation from the robot local coordinate system to the CCTV camera's local coordinate system. This transformation could be written as:

$$T_{Robot \to CCTV} = (T_{CCTV \to World})^{-1} \cdot T_{Robot \to World} = T_{World \to CCTV} \cdot T_{Robot \to World}$$

You could invert this transformation matrix $(T_{CCTV \to World})^{-1}$ using your favorite programming language (or linear algebra toolbox). However, as discussed in Task 1, the inverse of a rotation matrix is its transpose– and the inverse of a transformation matrix could be formulated as follows:

$$T = \begin{bmatrix} R_{3\times3} & t_{3\times1} \\ \mathbf{0}_{1\times3} & 1 \end{bmatrix}_{4\times4}$$

$$T^{-1} = \begin{bmatrix} R^T_{3\times3} & -R^T_{3\times3} * t_{3\times1} \\ \mathbf{0}_{1\times3} & 1 \end{bmatrix}_{4\times4}$$

In the same manner, we obtain the following:

$$T_{Robot \to CCTV} = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -4 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} & 0 & 6 \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 & 5 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} & 0 & 5 \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 & 5 \\ 0 & 0 & 1 & -4 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

- What is the neighbor's camera pose relative to the house? The neighbor's camera is translated 8.5m in the positive X-Axis direction, 7.5m in the negative Y-Axes directions, and 2m in the positive direction of the Z-Axis (in the world coordinate system). Like the robot, the object model for the camera is designed to look at the negative X-Axis direction (when no rotation is applied). In order for it to point at the house, we rotate it -45 degrees around the Z-Axis.

Answer The neighbor's camera relative to the house (i.e. pose of the house is the world coordinates itself) is:

$$T_{Neighbor \to World} = \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 & 8.5 \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 & -7.5 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

- What is the robot transformation (i.e. its pose) relative to the neighbor's camera?

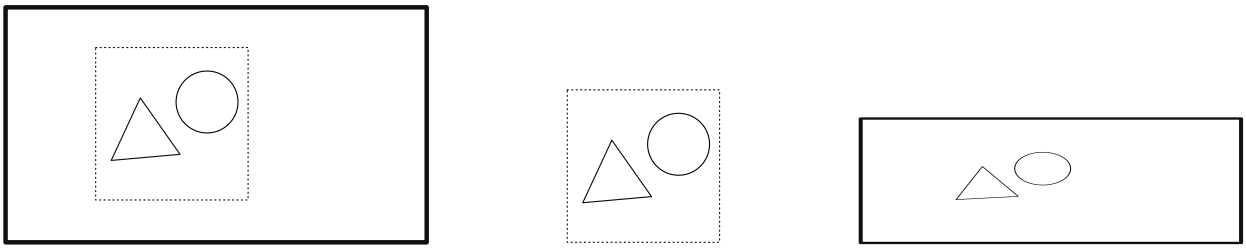
Answer

$$T_{World \to Neighbor} = (T_{Neighbor \to World})^{-1}$$

$$T_{Robot \to Neighbor} = T_{World \to Neighbor} * T_{Robot \to World} = \begin{bmatrix} 0 & -1 & 0 & -7.5\sqrt{2} \\ 1 & 0 & 0 & 5\sqrt{2} \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

From the above pose, we know that the robot is 2m lower than the neighbor's camera (as $z = -2$). We also know that the only rotation axis we have used so far is the Z-Axis. Therefore, the rotation enclosed in the above pose could be interpretted as a 90 degrees rotation around the Z-axis ($sin(\gamma) = 1$; Thus, $\gamma = \frac{\pi}{2}$). The robot coordinate system is rotated 90 degrees relative to the neghbor's camera coordinate system.

Task 3 Camera Intrinsics



a) b)

You are given an image of size (x: 3680, y: 2456) taken from a camera with known intrinsics matrix K.

$$K = \begin{bmatrix} 2960 & 0 & 1841 \\ 0 & 2960 & 1235 \\ 0 & 0 & 1 \end{bmatrix}$$

Find the new intrinsic matrix for the following cases:

a) A cropped image defined with the (width: 1000, height: 1000) and the center (x: 1500; y: 750) in respect of the original image.

Answer The camera intrinsic matrix is defined as follows:

$$K = \begin{bmatrix} f_x & \gamma & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

Where f_x and f_y represent the focal length in terms of pixels; c_x and c_y represent the principal point coordinates relative to the image plane origin; and γ represents the skew coefficient between the x and the y axis and is often 0.

The image plane coordinates system is usually defined with the origin in the top-left corner of the image, x axis pointing right and y axis pointing down.

The cropped image has it's own coordinates system with the origin in the top-left corner of the crop. In order to find the intrinsic matrix for the cropped image we need

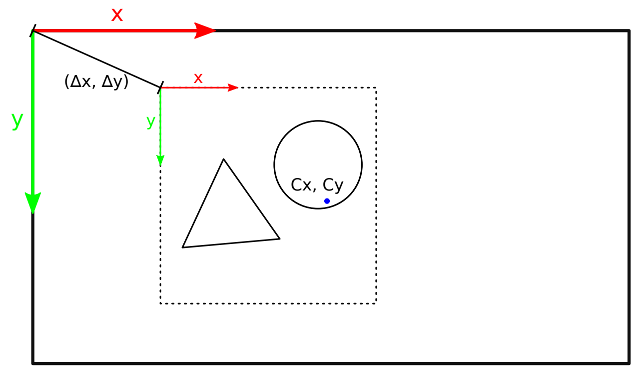


Figure 2

to estimate coordinates of the principal point of the original image (c_x^{full} and c_y^{full}) relative to the coordinates system of the cropped image (Figure 2).

Therefore the coordinates of the principal point of the cropped image can be found as follows:

$$c_x^{crop} = c_x^{full} - \underbrace{(p_x - W^{crop}/2)}_{\Delta_x} = 1841 - (1500 - 1000/2) = 841$$
$$c_y^{crop} = c_y^{full} - \underbrace{(p_y - H^{crop}/2)}_{\Delta_y} = 1235 - (750 - 1000/2) = 985$$

Where c_x^{full} and c_y^{full} are coordinates of the principal point of the original image; p_x and p_y are coordinates of the center of the crop in the coordinate system of the original image; W^{crop} and H^{crop} represent the size of the crop.

b) A downscaled image of size (x: 1840, y: 614)

Answer Image resizing can be represented with the 2D scaling transformation using homogeneous coordinates:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} S_x & 0 & 0 \\ 0 & S_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

Where S_x and S_y represent scale factors for x and y axes:

$$S_x = \frac{W^{resized}}{W^{original}} = \frac{1840}{3680} = 0.5$$
$$S_y = \frac{H^{resized}}{H^{original}} = \frac{614}{2456} = 0.25$$

To find the intrinsic matrix for the scaled image we need to apply this transformation to the original intrinsic matrix:

$$K^{resized} = \begin{bmatrix} 0.5 & 0 & 0 \\ 0 & 0.25 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2960 & 0 & 1841 \\ 0 & 2960 & 1235 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1480 & 0 & 920.5 \\ 0 & 740 & 308.75 \\ 0 & 0 & 1 \end{bmatrix}$$