Representing position and orientation

Jae Yun JUN KIM*

October 5, 2020

1 Introduction [1]

A fundamental requirement in robotics and computer vision is to represent the position and orientation of objects (e.g., robots, cameras, etc.) in an environment.

A **point in space** can be described by a **coordinate vector**. The **displacement of this point** with respect to some **reference coordinate frame** can be described by a vector. Such a reference coordinate frame is labelled as $\{A\}$, where A is an arbitrary name given to this frame. A **coordinate frame** (e.g., the **Cartesian coordinate system**) is a set of orthogonal axes which intersect at a point known as the **origin**.

We often consider not only a point in space but a set of points that comprise some object. We assume that the object is rigid and that its points maintain a constant relative position with respect to the object's coordinate frame. When we describe the **position** and the **orientation** of this object, we do so by describing the position and the orientation of its coordinate frame. A **coordinate frame** is labelled as $\{B\}$, where B is an arbitrary name given to this frame. Its **axes** are labelled as X_B and Y_B. The position and the orientation of a frame is known as its **pose**. The relative pose of a frame with respect to a reference coordinate frame is denoted by the symbol ξ. Hence, the relative pose of the frame $\{B\}$ with respect to the reference coordinate frame $\{A\}$ is denoted as ${}^A\xi_B$. Notice that if the initial superscript is missing in a notation, we can assume that it refers with respect to the world coordinate frame $\{O\}$.

The point $\mathbf{p}$ can be described with respect to either coordinate frame:

$$ {}^A\mathbf{p} = {}^A\xi_B \cdot {}^B\mathbf{p} \tag{1} $$

An important characteristic of **relative poses** is that they can be **composed** or **compounded**. If a frame can be described in terms of another by a relative pose, then they can be applied sequentially

$$ {}^A\xi_C = {}^A\xi_B \oplus {}^B\xi_C, \tag{2} $$

which means that the pose of C relative to A can be obtained by compounding the relative poses from A to B and B to C. $\oplus$ is an operator that indicates composition of relative poses. For this case, the point p can be described as

$$ {}^A\mathbf{p} = {}^A\xi_B \oplus {}^B\xi_C \cdot {}^C\mathbf{p} \tag{3} $$

So far, we worked with 2-dimensional coordinate frame, which may be appropriate for mobile robots that operate in a planar world. But, when we want to describe the pose of a flying robot,

*ECE Paris Graduate School of Engineering, 37 quai de Grenelle 75015 Paris, France; jae-yun.jun-kim@ece.fr

underwater robot or a manipulator, we required 3-dimensional coordinate frame. Suppose that there is a camera fixed to the ceiling of a room and that it observes a 3D object. Suppose that another camera mounted on a mobile robot at some constant relative pose also observes the same object. These two suppositions can be mathematically described as

$$\xi_F \oplus {}^F\xi_B = \xi_R \oplus {}^R\xi_C \oplus {}^C\xi_B. \tag{4}$$

2 Representing Pose in 2D [1]

A point is represented by its x- and y-coordinates (x, y)

$$\mathbf{p} = x\hat{\mathbf{x}} + y\hat{\mathbf{y}} \tag{5}$$

We can see that the origin of B is displaced by $t = (x, y)$ and then rotated counter-clockwisely by and angle θ. Hence,

$${}^A\xi_B \sim \begin{bmatrix} x \\ y \\ \theta \end{bmatrix}, \tag{6}$$

where $\sim$ indicates the equivalence. Unfortunately, this representation is not convenient for compounding since

$$(x_1, y_1, \theta_1) \oplus (x_2, y_2, \theta_2) \tag{7}$$

is a complex trigonometric function of both poses. Hence, we need to propose an appropriate representation for the change of poses.

$$\begin{aligned} \hat{\mathbf{x}}_\mathbf{B} &= \cos\theta\, \hat{\mathbf{x}}_\mathbf{V} + \sin\theta\, \hat{\mathbf{y}}_\mathbf{V} \\ \hat{\mathbf{y}}_\mathbf{B} &= -\sin\theta\, \hat{\mathbf{x}}_\mathbf{V} + \cos\theta\, \hat{\mathbf{y}}_\mathbf{V} \end{aligned} \tag{8}$$

which can be written in matrix form as

$$\begin{bmatrix} \hat{\mathbf{x}}_B & \hat{\mathbf{y}}_B \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{x}}_V & \hat{\mathbf{y}}_V \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{9}$$

We can represent the point p with respect to B as

$${}^Bp = {}^Bx\ \hat{x}_B + {}^By\ \hat{y}_B = \begin{bmatrix} \hat{\mathbf{x}}_\mathbf{B} & \hat{\mathbf{y}}_\mathbf{B} \end{bmatrix} \begin{bmatrix} {}^Bx \\ {}^By \end{bmatrix} \tag{10}$$

$${}^Bp = \begin{bmatrix} \hat{\mathbf{x}}_V & \hat{\mathbf{y}}_V \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} {}^Bx \\ {}^By \end{bmatrix} \tag{11}$$

$$\begin{bmatrix} {}^Vx \\ {}^Vy \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} {}^Bx \\ {}^By \end{bmatrix} \tag{12}$$

$$\begin{bmatrix} {}^Vx \\ {}^Vy \end{bmatrix} = {}^VR_B \begin{bmatrix} {}^Bx \\ {}^By \end{bmatrix} \tag{13}$$

The rotation matrix ${}^VR_B$ has some special properties.

- **orthonormal** since each of its columns is a unit vector and the columns are orthogonal. The columns are the unit vectors that define B with respect to V

- The determinant is +1, which means that R belongs to the **special orthogonal group** of dimension 2 or $R \in SO(S) \in \mathbb{R}^{2\times 2}$. The unit determinant means that the length of a vector is unchanged after transformation, that is, $\left|{}^B\mathbf{p}\right| = \left|{}^V\mathbf{p}\right|, \forall\theta$.

- Orthonormal matrices have the property $\mathbf{R}^{-1} = \mathbf{R}^T$. That is, the inverse is the same as the transpose

$$\begin{bmatrix} {}^B x \\ {}^B y \end{bmatrix} = \left({}^V R_B\right)^{-1} \begin{bmatrix} {}^V x \\ {}^V y \end{bmatrix} = \left({}^V R_B\right)^T \begin{bmatrix} {}^V x \\ {}^V y \end{bmatrix} = {}^B R_V \begin{bmatrix} {}^V x \\ {}^V y \end{bmatrix} \tag{14}$$

  Notice that investing the matrix is the same as swapping the superscript and subscript, which leads to the identity $R(-\theta) = R(\theta)^T$.

The second part of representing pose is to account for the translation between the origins of the frames

$$\begin{bmatrix} {}^A x \\ {}^A y \end{bmatrix} = \begin{bmatrix} {}^V x \\ {}^V y \end{bmatrix} + \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} {}^B x \\ {}^B y \end{bmatrix} + \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta & x \\ \sin\theta & \cos\theta & y \end{bmatrix} \begin{bmatrix} {}^B x \\ {}^B y \\ 1 \end{bmatrix} \tag{15}$$

We can express the above expression more compactly as

$$\begin{bmatrix} {}^A x \\ {}^A y \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{{}^A R_B} & \mathbf{t} \\ \mathbf{0_{1\times 2}} & 1 \end{bmatrix} \begin{bmatrix} {}^B x \\ {}^B y \\ 1 \end{bmatrix}, \tag{16}$$

where $\mathbf{t} = (x, y)^T$ is the translation of the frame. Notice that $\mathbf{{}^A R_B} = \mathbf{{}^V R_B}$ since the axes of $\{A\}$ and $\{B\}$ are parallel. The above expression is known as the **homogenenous transformation** and can be rewritten as

$${}^A\tilde{\mathbf{p}} = \begin{bmatrix} \mathbf{{}^A R_B} & \mathbf{t} \\ \mathbf{0_{1\times 2}} & 1 \end{bmatrix} {}^B\tilde{\mathbf{p}} = \mathbf{{}^A T_B} {}^B\tilde{\mathbf{p}}, \tag{17}$$

where $\mathbf{{}^A T_B}$ is referred to as a **homogeneous transformation**, and ${}^B\tilde{\mathbf{p}} = [x, y, 1]^T$. This matrix has a specific structure and belongs to the **spacial Euclidean Group** of dimension 2. That is, $\mathbf{T} \in SE(2) \subset \mathbb{R}^{3\times 3}$.

We now observe that

$$\xi(x, y, \theta) \sim \begin{bmatrix} \cos\theta & -\sin\theta & x \\ \sin\theta & \cos\theta & y \\ 0 & 0 & 1 \end{bmatrix} \tag{18}$$

3 Representing pose in 3-dimensions [1]

The 3-dimensional case is an extended version of the 2-dimensional case by adding a new coordinate axis denoted by z. z-axis

- is orthogonal to both the x- and y- axes.

- obeys the **right-hand rule** and forms a **right-handed coordinate frame**.

A **right-handed coordinate frame** is defined by the first three fingers of your right hand which indicate the relative directions of the x-, y- and z-axes, respectively (**Right-hand rule**).

Unit vectors parallel to the axes are denoted

$$\begin{aligned}\hat{z} &= \hat{\mathbf{x}} \times \hat{\mathbf{y}} \\ \hat{x} &= \hat{\mathbf{y}} \times \hat{\mathbf{z}} \\ \hat{y} &= \hat{\mathbf{z}} \times \hat{\mathbf{x}}\end{aligned} \tag{19}$$

A point $\mathbf{p}$ is represented by its x-, y-, and z- coordinates

$$\mathbf{p} = x\hat{\mathbf{x}} + y\hat{\mathbf{y}} + z\hat{\mathbf{z}} \tag{20}$$

3.1 Representing orientation in 3-dimensions

Euler's rotation theorem (Kuipers 1999): Any two independent orthonormal coordiante frames can be related by a sequence of rotations (no more than three) about coordina axes, where no two successive rotations may be about the same axis.

We observe that a sequence of two rotations in different orders result in different final orientations.

3.1.1 Orthonormal rotation matrix

As for the 2-dimensional case, we can represent the orientation of a coordinate frame by its unit vectors in terms of the reference coordinate frame. Each unit vector has three elements and they form the columns of a 3×3 **orthonormal matrix** ${}^{A}\mathbf{R}_{B}$

$$\begin{bmatrix} {}^{A}X \\ {}^{A}Y \\ {}^{A}Z \end{bmatrix} = {}^{A}\mathbf{R}_{B} \begin{bmatrix} {}^{B}X \\ {}^{B}Y \\ {}^{B}Z \end{bmatrix} \tag{21}$$

which rotates a vector defined with respect to frame $\{B\}$ to a vector with respect to $\{A\}$. The matrix $\mathbf{R}$ belongs to the special orthogonal group of dimension 3 or $\mathbf{R} \in SO(3) \subset \mathbf{R}^{3\times3}$. Its properties are that $\mathbf{R}^T = \mathbf{R}^{-1}$ and $\det(\mathbf{R}) = 1$.

The orthonormal rotation matrices for rotation of θ about the x-, y-, and z-axes are

$$R_x(\theta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix}, R_y(\theta) = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix}, R_z(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{22}$$

3.1.2 Three-angle representations

Euler's rotation theorem successive rotation about three axes such that no two successive rotation are about the same axis. There are two classes of rotation sequence: **Eulerian** and **Cardanian**, named after **Euler** and **Cardano**, respectively.

- The **Eulerian** type involves repetition, but not successive, of rotations about one particular axis: XYX, XZX, YXY, YZY, ZXZ, or ZYZ.
- The **Cardanian** type is characterized by rotations about all three axes: XYZ, XZY, YZX, YXZ, ZXY, or ZYX.

A widely used three-angle representation is the **roll-pitch-yaw** angle sequence

$$\mathbf{R} = \mathbf{R_x}(\theta_r)\mathbf{R_y}(\theta_p)\mathbf{R_z}(\theta_y) \tag{23}$$

The roll-pitch-yaw sequence allows all angles to have arbitrary sign, and it has a singularity when $\theta_p = \pm\frac{\pi}{2}$, which is fortunately outside the range of feasible attitudes for most vehicles.

3.2 Combining translation and orientation

We can use a **homogeneous transformation matrix** to describe rotation and translation

$$\begin{bmatrix} {}^{A}x \\ {}^{A}y \\ {}^{A}z \\ 1 \end{bmatrix} = \begin{bmatrix} {}^{A}\mathbf{R}_B & \mathbf{t} \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix} \begin{bmatrix} {}^{B}x \\ {}^{B}y \\ {}^{B}z \\ 1 \end{bmatrix} \tag{24}$$

When we concatenate two transformations, we have

$$\mathbf{T}_1\mathbf{T}_2 = \begin{bmatrix} \mathbf{R}_1 & \mathbf{t}_1 \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{R}_2 & \mathbf{t}_2 \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R}_1\mathbf{R}_2 & \mathbf{t}_1 + \mathbf{R}_1\mathbf{t}_2 \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix} \tag{25}$$

On the other hand, we observe that

$$\mathbf{T}^{-1} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{R}^T & -\mathbf{R}^T\mathbf{t} \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix} \tag{26}$$

4 Denavit-Hartenberg parameters [2]

There are four Denavit-Hartenberg parameters:

- a_i: the distance from $\hat{Z}_i$ to $\hat{Z}_{i+1}$ measured along $\hat{X}_i$;
- α_i: the angle from $\hat{Z}_i$ to $\hat{Z}_{i+1}$ measured about $\hat{X}_i$;
- d_i: the distance from $\hat{X}_{i-1}$ to $\hat{X}_i$ measured along $\hat{Z}_i$;
- θ_i: the angle from $\hat{X}_{i-1}$ to $\hat{X}_i$ measured along $\hat{Z}_i$.

References

[1] Peter Corke, Robotics, vision and control, Springer-Verlag Berlin Heidelberg, 2011.

[2] John J. Craig, Introduction to robotics: mechanics and control, Pearson, 3rd edition, 2004