

¹ Chapter 4

² Rigid-body transforms and ³ mapping

Reading

1. LaValle Chapter 3.2 for rotation matrices, Chapter 4.1-4.2 for quaternions
2. Thrun Chapter 9.1-9.2 for occupancy grids
3. OctoMap: An Efficient Probabilistic 3D Mapping Framework Based on Octrees
<http://www.arminhornung.de/Research/pub/hornung13auro.pdf>,
also see <https://octomap.github.io>.
4. Robot Operating System
<http://www.willowgarage.com/sites/default/files/icraoss09-ROS.pdf>, Optional: Lightweight Communications and Marshalling (LCM) system
<https://people.csail.mit.edu/albert/pubs/2010-huang-olson-moore-lcm-iros.pdf>
5. A Perception-Driven Autonomous Urban Vehicle
<https://april.eecs.umich.edu/media/pdfs/mitduc2009.pdf>
6. Optional reading: Thrun Chapter 10 for simultaneous localization and mapping

⁴ In the previous chapter, we looked at ways to estimate the state of the robot
⁵ in the physical world. We kept our formulation abstract, e.g., the way the robot
⁶ moves was captured by an abstract expression like $x_{k+1} = f(x_k, u_k) + \epsilon$
⁷ and observations $y_k = g(x_k) + \nu$ were similarly opaque. In order to actually
⁸ implement state estimation algorithms on real robots, we need to put concrete
⁹ functions in place of f, g .

1 This is easy to do for some robots, e.g., the robot in Problem 1 in Home-
 2 work 1 moved across cells. Of course real robots are a bit more complicated,
 3 e.g., a car cannot move sideways (which is a huge headache when you par-
 4 allel park). In the first half of this chapter, we will look at how to model the
 5 dynamics f using rigid-body transforms.

6 The story of measurement models and sensors is similar. Although we
 7 need to write explicit formulae in place of the abstract function g . In the
 8 second half, we will study occupancy grids and dig deeper into a typical
 9 state-estimation problem in robotics, namely that of mapping the location of
 10 objects in the world around the robot.

11 4.1 Rigid-Body Transformations

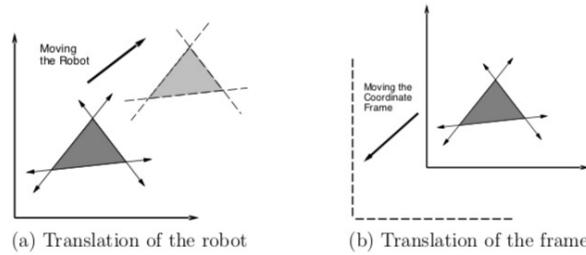
12 Let us imagine that the robot has a rigid body, we think of this as a subset
 13 $A \subset \mathbb{R}^2$. Say the robot is a disc

$$A = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}.$$

14 This set A changes as the robot moves around, e.g., if the center of mass of
 15 the robot is translated by $x_t, y_t \in \mathbb{R}$ the set A changes to

$$A' = \{(x + x_t, y + y_t) : (x, y) \in A\}.$$

16 The concept of “degrees of freedom” denotes the maximum number of in-
 17 dependent parameters needed to completely characterize the transformation
 18 applied to a robot. Since the set of allowed values (x_t, y_t) is a two-dimensional
 19 subset of \mathbb{R}^2 , then the degrees of freedom available to a translating robot is
 20 two.



21

22 As the above figure shows, there are two ways of thinking about this transfor-
 23 mation. We can either think of the robot transforming while the co-ordinate
 24 frame of the world is fixed, or we can think of it as the robot remaining sta-
 25 tionary and the co-ordinate frame undergoing a translation. The second style
 26 is useful if you want to imagine things from the robot’s perspective. But the
 27 first one feels much more natural and we will therefore exclusively use the
 28 first notion.

29 If the same robot if it were rotated counterclockwise by some angle
 30 $\theta \in [0, 2\pi]$, we would map

$$(x, y) \mapsto (x \cos \theta - y \sin \theta, x \sin \theta + y \cos \theta).$$

¹ Such a map can be written as multiplication by a 2×2 rotation matrix

$$R(\theta) = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}. \quad (4.1)$$

² to get

$$\begin{bmatrix} x \cos \theta - y \sin \theta \\ x \sin \theta + y \cos \theta \end{bmatrix} = R(\theta) \begin{bmatrix} x \\ y \end{bmatrix}.$$

³ The transformed robot is thus given by

$$A' = \left\{ R \begin{bmatrix} x \\ y \end{bmatrix} : (x, y) \in A \right\}.$$

⁴ If we perform both rotation and translation, we can the transformation using a
⁵ single matrix

$$T = \begin{bmatrix} \cos \theta & -\sin \theta & x_t \\ \sin \theta & \cos \theta & y_t \\ 0 & 0 & 1 \end{bmatrix} \quad (4.2)$$

⁶ and this transformation looks like

$$\begin{bmatrix} x \cos \theta - y \sin \theta + x_t \\ x \sin \theta + y \cos \theta + y_t \\ 1 \end{bmatrix} = T \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}.$$

⁷ The point $(x, y, 1) \in \mathbb{R}^3$ is called homogeneous coordinate space correspond-
⁸ ing to $(x, y) \in \mathbb{R}^2$ and the matrix T is called a *homogeneous transformation*
⁹ *matrix*. The peculiar names comes from the fact that even if the matrix T
¹⁰ maps rotations and translations of rigid bodies $A \subset \mathbb{R}^2$, it is just a linear
¹¹ transformation of the point $(x, y, 1)$ if viewed in the larger space \mathbb{R}^3 .

¹² **Rigid-body transformations** The transformations $R \in \mathbb{R}^{2 \times 2}$ or $T \in \mathbb{R}^{3 \times 3}$
¹³ are called rigid-body transformations. Mathematically, it means that they do
¹⁴ not cause the distance between any two points inside the set A to change. Rigid-
¹⁵ body transformations are what are called an orthogonal group in mathematics.

¹⁶ **A group** is a mathematical object which imposes certain conditions upon
¹⁷ how two operations, e.g., rotations, can be composed together. For instance,
¹⁸ if G is the group of rotations, then (i) the composition of two rotations is a
¹⁹ rotation, we say that it satisfies closure $R(\theta_1)R(\theta_2) \in G$, (ii) rotations are
²⁰ associative

$$R(\theta_1) \{R(\theta_2)R(\theta_3)\} = \{R(\theta_1)R(\theta_2)\} R(\theta_3),$$

²¹ and, (iii) there exists an identity and inverse rotation

$$R(0), R(-\theta) \in G.$$

²² **An orthogonal group** is a group whose operations preserve distances in
²³ Euclidean space, i.e., $g \in G$ is an element of the group that acts on two points

❶ It is important to remember that T represents rotation *followed* by a translation, not the other way around.

¹ $x, y \in \mathbb{R}^d$ then

$$\|g(x) - g(y)\| = \|x - y\|.$$

² If we identify the basis in Euclidean space to be the set of orthonormal vectors $\{e_1, \dots, e_d\}$, then equivalently, the orthogonal group $O(d)$ is the set of ³ orthogonal matrices ⁴

$$O(d) := \{O \in \mathbb{R}^{d \times d} : OO^\top = O^\top O = I\}.$$

⁵ This implies that the square of the determinant of any element $a \in O(d)$ is 1, ⁶ i.e., $\det(a) = \pm 1$.

❶ Check that any rotation matrix R belongs to an orthogonal group.

⁷ **The Special Orthogonal Group** is a sub-group of the orthogonal group ⁸ where the determinant of each element is +1. You can see that rotations are a ⁹ special orthogonal group. We denote rotations of objects in \mathbb{R}^2 as

$$SO(2) := \{R \in \mathbb{R}^{2 \times 2} : R^\top R = RR^\top = I, \det(R) = 1\}. \quad (4.3)$$

¹⁰ Each group element $g \in SO(2)$ denotes a rotation of the XY -plane about the ¹¹ Z -axis. The group of 3D rotations is called the Special Orthogonal Group ¹² $SO(3)$ and is defined similarly

$$SO(3) := \{R \in \mathbb{R}^{3 \times 3} : R^\top R = RR^\top = I, \det(R) = 1\}. \quad (4.4)$$

¹³ **The Special Euclidean Group** $SE(2)$ is simply a composition of a 2D ¹⁴ rotation $R \in SO(2)$ and a 2D translation $\mathbb{R}^2 \ni v \equiv (x_t, y_t)$

$$SE(2) = \left\{ \begin{bmatrix} R & v \\ 0 & 1 \end{bmatrix} : R \in SO(2), v \in \mathbb{R}^2 \right\} \subset \mathbb{R}^{3 \times 3}. \quad (4.5)$$

¹⁵ The Special Euclidean Group $SE(3)$ is defined similarly as

$$SE(3) = \left\{ \begin{bmatrix} R & v \\ 0 & 1 \end{bmatrix} : R \in SO(3), v \in \mathbb{R}^3 \right\} \subset \mathbb{R}^{4 \times 4}; \quad (4.6)$$

¹⁶ again, remember that it is rotation *followed* by a translation.

¹⁷ 4.1.1 3D transformations

¹⁸ Translations and rotations in 3D are conceptually similar to the two-dimensional ¹⁹ case; however the details appear a bit more difficult because rotations in 3D ²⁰ are more complicated.

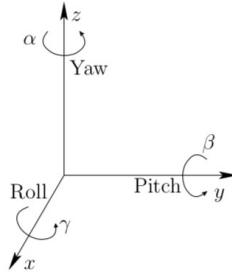


Figure 4.1: Any three-dimensional rotation can be described as a sequence of rotations about each of the cardinal axes. We usually give these specific names: rotation about the Z -axis is called yaw, rotation about the X -axis is called roll and rotation about Y -axis is called pitch. You should commit this picture and these names to memory because it will be of enormous to think about these rotations intuitively.

- ¹ **Euler angles** We know that a pure counter-clockwise rotation about one of the axes is written in terms of a matrix, say yaw of α -radians about the Z -axis

$$R_z(\alpha) = \begin{bmatrix} \cos \alpha & -\sin \alpha & 0 \\ \sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

- ³ Notice that this is a 3×3 matrix that keeps the Z -coordinate unchanged and ⁴ only affects the other two coordinates. Similarly we have for pitch (β about the Y -axis) and roll (γ about the X -axis)

$$R_y(\beta) = \begin{bmatrix} \cos \beta & 0 & \sin \beta \\ 0 & 1 & 0 \\ -\sin \beta & 0 & \cos \beta \end{bmatrix}, R_x(\gamma) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \gamma & -\sin \gamma \\ 0 & \sin \gamma & \cos \gamma \end{bmatrix}.$$

- ⁶ By convention, a rotation matrix in three dimensions is understood as a sequential application of rotations, *first roll, then pitch, and then yaw*

$$\mathbb{R}^{3 \times 3} = R(\alpha, \beta, \gamma) = R_z(\alpha)R_y(\beta)R_x(\gamma). \quad (4.7)$$

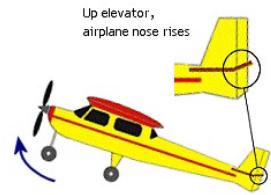
- ⁸ The angles (α, γ, β) are called Euler angles. Imagine how the body frame of ⁹ the robot changes as successive rotations are applied. If you were sitting in a ¹⁰ car, a pure yaw would be similar to the car turning left; the Z -axis corresponding to this yaw would however only be pointing straight up perpendicular to ¹¹ the ground if you had not performed a roll/pitch before. If you had, the Z -axis ¹² of the body frame with respect to the world will be tilted.

- ¹⁴ Another important thing to note is that one single parameter determines all ¹⁵ possible rotations about one axis, i.e., $SO(2)$. But three Euler angles are used ¹⁶ to parameterize general rotations in three-dimensions.

❶ Here is how I remember these names. Say you are driving a car, usually in robotics we take the X -axis to be longitudinally forward, the Y -axis is your left hand if you are in the driver's seat and the Z -axis points up by the right-hand thumb rule. Roll



is what a dog does when it rolls, it rotates about the X -axis. Pitch is what a plane



does when it takes off, its nose lifts up and it rotates about the Y -axis. Yaw is the one leftover.

¹ **Rotation matrices to Euler angles** We can back-calculate the Euler angles
² from a rotation matrix as follows. Given an arbitrary matrix

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix},$$

³ we set

$$\begin{aligned} \alpha &= \tan^{-1}(r_{21}/r_{11}) \\ \beta &= \tan^{-1}\left(-r_{31}/\sqrt{r_{32}^2 + r_{33}^2}\right) \\ \gamma &= \tan^{-1}(r_{32}/r_{33}). \end{aligned} \quad (4.8)$$

⁴ For each angle, the corresponding quadrant for the Euler angle is determined
⁵ using the signs of the numerator and the denominator. So you should use
⁶ the function atan2 in Python/C++ to implement these expressions correctly.
⁷ Notice that some of the expressions have r_{11} and r_{33} in the denominator, this
⁸ means that we need $r_{11} = \cos \alpha \cos \beta = 0$ and $r_{33} = \cos \beta \cos \gamma = 0$. Euler
⁹ angles do not completely parametrize the group of rotations $SO(3)$. This is
¹⁰ really an indicator that this particular physical rotation can be parameterized
¹¹ in two different ways using Euler angles, so the map from rotation matrices to
¹² Euler angles is not unique.

¹³ **Homogeneous coordinates in three dimensions** Just like the 2D case, we
¹⁴ can define a 4×4 matrix that transforms points $(x, y, z) \in \mathbb{R}^3$ to their new
¹⁵ locations after a rotation by Euler angles (α, β, γ) and a translation by a vector
¹⁶ $v = (x_t, y_t, z_t) \in \mathbb{R}^3$

$$T = \begin{bmatrix} R(\alpha, \beta, \gamma) & v \\ 0 & 1 \end{bmatrix}.$$

¹⁷ 4.1.2 Rodrigues' formula: an alternate view of rotations

¹⁸ Consider a point $r(t) \in \mathbb{R}^3$ that is being rotated about an axis denoted by a
¹⁹ unit vector $\omega \in \mathbb{R}^3$ with an angular velocity of 1 radian/sec. The instantaneous
²⁰ linear velocity of the head of the vector is

$$\dot{r}(t) = \omega \times r(t) \equiv \hat{\omega}r(t) \quad (4.9)$$

²¹ where the \times denotes the cross-product of the two vectors $a, b \in \mathbb{R}^3$

$$a \times b = \begin{bmatrix} a_2 b_3 - a_3 b_2 \\ a_3 b_1 - a_1 b_3 \\ a_1 b_2 - a_2 b_1 \end{bmatrix}$$

²² which we can equivalently denote as a matrix vector multiplication $a \times b = \hat{a}b$
²³ where

$$\hat{a} = \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix}. \quad (4.10)$$

1 is a skew-symmetric matrix. The solution of the differential equation (4.9) at
 2 time $t = \theta$ is

$$r(\theta) = \exp(\hat{\omega}\theta) r(0)$$

3 where the matrix exponential of a matrix A is defined as

$$\exp(A) = I + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \dots = \sum_{k=0}^{\infty} \frac{A^k}{k!}.$$

4 This is an interesting observation: a rotation about a fixed axis ω by an angle θ
 5 can be represented by the matrix

$$R = \exp(\hat{\omega}\theta).$$

6 You can check that this matrix is indeed a rotation by showing that $R^\top R = I$
 7 and that $\det(R) = +1$. We can expand the matrix exponential and collect odd
 8 and even powers of $\hat{\omega}$ to get

$$R = I + \sin \theta \hat{\omega} + (1 - \cos \theta) \hat{\omega}^2. \quad (4.11)$$

9 which is the Rodrigues' formula that relates the angle θ and the axis ω to the
 10 rotation matrix. We can also go in the opposite direction, i.e., given a matrix
 11 R calculate what angle θ and axis ω it corresponds to using

$$\begin{aligned} \cos \theta &= \frac{\text{tr}(R) - 1}{2} \\ \hat{\omega} &= \frac{R - R^\top}{2 \sin \theta}. \end{aligned} \quad (4.12)$$

12 4.2 Quaternions

13 We know two ways to think about rotations: we can either think in terms of the
 14 three Euler angles (α, β, γ) , or we can consider a rotation matrix $R \in \mathbb{R}^{3 \times 3}$.
 15 We also know ways to go to and fro between these two forms with the caveat
 16 that solving for Euler angles using (4.8) may be degenerate in some cases.
 17 While rotation matrices are the most general representation of rotations, using
 18 them in computer code is cumbersome (it is, after all, a matrix of 9 elements),
 19 imagine an EKF where the state is a rotation matrix. You can certainly
 20 implement the same filter using Euler angles but doing so requires special
 21 care due to the degeneracies. Quaternions are a neat way to avoid both these
 22 problems, they parametrize the space of rotations using 4 numbers and do not
 23 have degeneracies.

24 The central idea behind quaternions is Euler's theorem which says that
 25 any 3D rotation can be considered as a pure rotation by an angle $\theta \in \mathbb{R}$ about
 26 an axis given by the unit vector ω . This is the result that we also exploited in
 27 Rodrigues' formula.

❶ Quaternions were invented by British mathematician William Rowan Hamilton while walking along a bridge with his wife. He was quite excited by this discovery and promptly graffitied the expression into the stone of the bridge

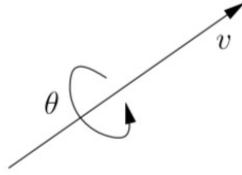


Figure 4.2: Any rotation in 3D can be represented using a unit vector ω and an angle $\theta \in \mathbb{R}$. Notice that there are two ways to encode the same rotation, the unit vector $-\omega$ and angle $2\pi - \theta$ would give the same rotation. Mathematicians say this as quaternions being a double-cover of $SO(3)$.

1 A quaternion q as a four-dimensional vector $q \equiv (u_0, u_1, u_2, u_3)$ and we
2 write it as

$$\begin{aligned} q &\equiv (u_0, u), \text{ or} \\ q &= u_0 + u_1 i + u_2 j + u_3 k, \end{aligned} \tag{4.13}$$

3 with i, j, k being three “imaginary” components of the quaternion with “complex-
4 numbers like” relationships

$$i^2 = j^2 = k^2 = ijk = -1. \tag{4.14}$$

5 It follows from these relationships that

$$ij = -ji = k, ki = -ik = j, \text{ and } jk = -jk = i.$$

6 Although you may be tempted to think about this, these imaginary components
7 i, j, k have no relationship with the square roots of negative unity used to
8 define standard complex numbers. You should simply think of the quaternion
9 as a four-dimensional vector in some space. A unit quaternion, i.e., one with

$$u_0^2 + u_1^2 + u_2^2 + u_3^2 = 1,$$

10 is special: unit quaternions can be used to represent rotations in 3D.

11 **Quaternion to axis-angle representation** The quaternion $q = (u_0, u)$ cor-
12 responds to a counter-clockwise rotation of angle θ about a unit vector ω
13 where θ and ω are such that

$$u_0 = \cos \frac{\theta}{2}, \text{ and } u = \sin \frac{\theta}{2} \omega. \tag{4.15}$$

14 So given an axis-angle representation of rotation like in Rodrigues’ formula
15 (θ, ω) we can write the quaternion as

$$q = \left(\cos \frac{\theta}{2}, \sin \frac{\theta}{2} \omega \right).$$

-
- ¹ Using this, we can also compute the inverse of a quaternion (rotation of angle
² θ about the opposite axis $-\omega$) as

$$q^{-1} := \left(\cos \frac{\theta}{2}, -\sin \frac{\theta}{2} \omega \right).$$

- ³ The inverse quaternion is therefore the quaternion where all entries except the
⁴ first have their signs flipped.

- ⁵ **Multiplication of quaternions** Just like two rotation matrices multiply
⁶ together to give a new rotation (rotations are a group), since quaternions are
⁷ also a representation for the group of rotations, we can also multiply two
⁸ quaternions $q_1 = (u_0, u)$, $q_2 = (v_0, v)$ together using the quaternion identities
⁹ for i, j, k in (4.14) to get a new quaternion

$$q_1 q_2 \equiv (u_0, u) \cdot (v_0, v) = (u_0 v_0 - u^\top v, u_0 v + v_0 u + u \times v).$$

¹⁰

❶ Quaternions belong to a larger group than rotations called the Symplectic Group $Sp(1)$.

- ¹¹ **Pure quaternions** A pure quaternion is a quaternion with a zero scalar value
¹² $u_0 = 0$. This is very useful to simply store a standard 3D vector $u \in \mathbb{R}^3$ as a
¹³ quaternion $(0, u)$. We can then rotate points easily between different frames
¹⁴ as follows. Given a vector $x \in \mathbb{R}^3$ we can form a quaternion $(0, x)$ and show
¹⁵ that

$$q \cdot (0, x) \cdot q^* = (0, R(q)x). \quad (4.16)$$

- ¹⁶ where $q^* = (u_0, -u)$ is the conjugate quaternion of $q = (u_0, u)$; the conjugate
¹⁷ is the same as the inverse for unit quaternions. Notice how the right-hand side
¹⁸ is the vector $R(q)x$ corresponding to the vector x rotation by a matrix $R(q)$.

- ¹⁹ **Quaternions to rotation matrix** The rotation matrix corresponding to a
²⁰ quaternion is

$$\begin{aligned} R(q) &= (u_0^2 - u^\top u) I_{3 \times 3} + 2 \frac{u_0 u}{\|u\|} + 2 u u^\top \\ &= \begin{bmatrix} 2(u_0^2 + u_1^2) - 1 & 2(u_1 u_2 - u_0 u_3) & 2(u_1 u_3 + u_0 u_2) \\ 2(u_1 u_2 + u_0 u_3) & 2(u_0^2 + u_2^2) - 1 & 2(u_2 u_3 - u_0 u_1) \\ 2(u_1 u_3 - u_0 u_2) & 2(u_2 u_3 + u_0 u_1) & 2(u_0^2 + u_3^2) - 1 \end{bmatrix}. \end{aligned} \quad (4.17)$$

- ²¹ Using this you can show the identity that rotation matrix corresponding to the
²² product of two quaternions is the product of the individual rotation matrices

$$R(q_1 q_2) = R(q_1) R(q_2).$$

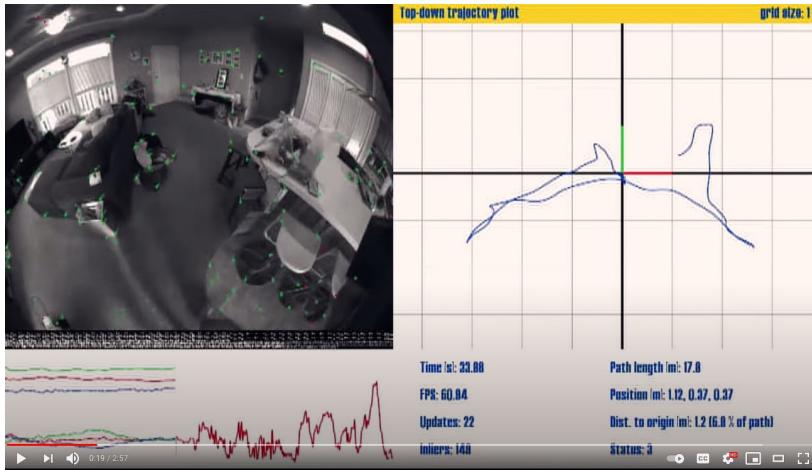
- ¹ **Rotation matrix to quaternion** We can also go in the reverse direction.
² Given a rotation matrix R , the quaternion is

$$\begin{aligned}
 u_0 &= \frac{1}{2} \sqrt{r_{11} + r_{22} + r_{33} + 1} \\
 \text{if } u_0 \neq 0, \quad u_1 &= \frac{r_{32} - r_{23}}{4u_0} \\
 u_2 &= \frac{r_{13} - r_{31}}{4u_0} \\
 u_3 &= \frac{r_{21} - r_{12}}{4u_0} \\
 \text{if } u_0 = 0, \quad u_1 &= \frac{r_{13}r_{12}}{\sqrt{r_{12}^2r_{13}^2 + r_{12}^2r_{23}^2 + r_{13}^2r_{23}^2}} \\
 u_2 &= \frac{r_{12}r_{23}}{\sqrt{r_{12}^2r_{13}^2 + r_{12}^2r_{23}^2 + r_{13}^2r_{23}^2}} \\
 u_3 &= \frac{r_{13}r_{23}}{\sqrt{r_{12}^2r_{13}^2 + r_{12}^2r_{23}^2 + r_{13}^2r_{23}^2}}.
 \end{aligned} \tag{4.18}$$

³ 4.3 Occupancy Grids

- ⁴ Rotation matrices and quaternions let us capture the dynamics of a rigid robot
⁵ body. We will next look at how to better understand observations.

- ⁶ **What is location and what is mapping?** Imagine a robot that is moving
⁷ around in a house. A natural representation of the state of this robot is the 3D lo-
⁸ cation of all the interesting objects in the room, e.g., <https://www.youtube.com/watch?v=Qe10ExwzCqk>.
⁹ At each time-instant, we record an observation from our sensor (in this case,
¹⁰ a camera) that indicates how far an object is from the robot. This helps us
¹¹ discover the location of the objects in the room. After gathering enough ob-
¹² servations, we would have created a *map* of the entire house. This map is the
¹³ set of positions of all interesting objects in the room. Such a map is called a
¹⁴ “feature map”, these are all the green points in the image below



- ¹⁶ The main point to understand about feature map is that we can hand over
¹⁷ this map to another robot that comes to the same house. The robot compares

1 images from its camera and if it finds one of the objects inside the map, it
 2 can get an estimate of its location/orientation in the room with respect to the
 3 known location of the object in the map. The map is just a set of “features”
 4 that help identify salient objects in the room (objects which can be easily
 5 detected in images and relatively uniquely determine the location inside the
 6 room). The second robot using this map to estimate its position/orientation in
 7 the room is called the *localization problem*. We already know how to solve
 8 the localization problem using filtering.

9 The first robot was solving a harder problem called *Simultaneous Local-
 10 ization And Mapping (SLAM)*: namely that of discovering the location of both
 11 itself and the objects in the house. This is a very important and challenging
 12 problem in robots but we will not discuss it further. MEAM 620 digs deeper
 13 into it.

In this section, we will solve a part of the SLAM problem, namely the mapping problem. We will assume that we know the position/orientation of the robot in the 3D world, and want to build a map of the objects in the world. We will discuss grid maps, which are a more crude way of representing maps than feature maps but can be used easily even if there are lots of objects.

14 **Grid maps** We will first discuss two-dimensional grid maps, they look as
 15 follows.

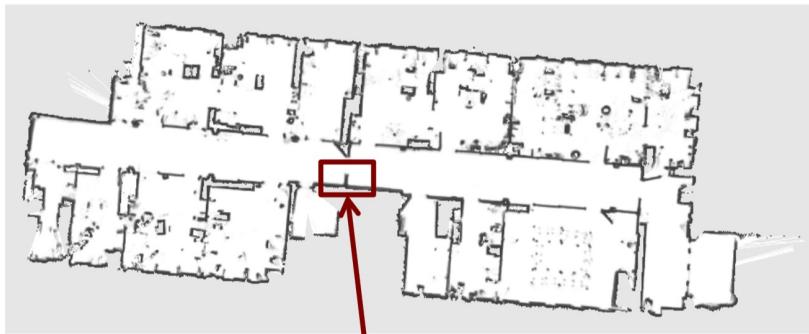
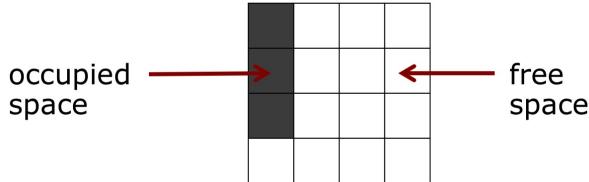


Figure 4.3: A grid map (also called an occupancy grid) is a large gray-scale image, each pixel represents a cell in the physical world. In this picture, cells that are occupied are colored black and empty cells represent free space. A grid map is a useful representation for a robot to localize in this house using observations from its sensors and comparing those to the map.

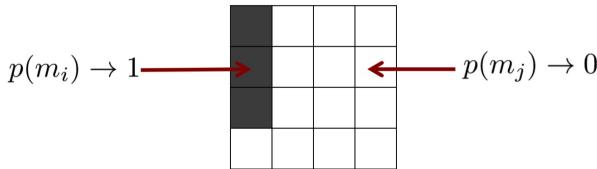
16 To get a quick idea of what we want to do, you can watch the mapping
 17 being performed in <https://www.youtube.com/watch?v=JJhEkIA1xSE>. We are
 18 interested in learning such maps from the observations that a robot collects as
 19 it moves around the physical space. Let us make two simplifying assumptions.

20 **Assumption 1: each cell is either free or occupied**



1

- 2 This is neat: we can now model each cell as a binary random variable that
3 indicates occupancy. Let the probability that the cell m_i be occupied be $p(m_i)$

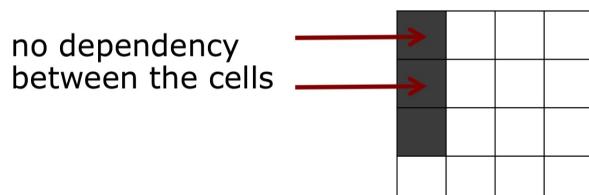


4

- 5 If we have $p(m_i) = 0$, then the cell is not occupied and if we have $p(m_i) = 1$,
6 then the cell is occupied. A priori, we do not know the state of the cell so we
7 will set the prior probability to be $p(m_i) = 0.5$.

- 8 **Assumption 2: the world is static** Objects in the world do not move. This
9 is reasonable if we are interested in estimating in building a map of the walls
10 inside the room. Note that it is not a reasonable assumption if there are moving
11 people inside the room. We will see a clever hack where the Bayes rule helps
12 automatically disregard such moving objects in this section.

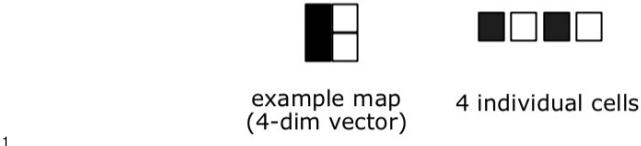
- 13 **Assumption 3: cells are independent of each other** This is another drastic
14 simplification. The state of our system is the occupancy of each cell in the
15 grid map. We assume that before receiving any observations, the occupancy of
16 each individual cell is independent; it is a Bernoulli variable with probability
17 1/2 since we have assumed the prior to be uniform in Assumption 1.



18

- 19 This means that if cells in the map are denoted by a vector $m = (m_1, \dots,)$,
20 then the probability of the cells being occupied/not-occupied can be written as

$$21 p(m) = \prod_i p(m_i). \quad (4.19)$$



1

2 4.3.1 Estimating the map from the data

- 3 Say that the robot pose (position and orientation) is given by the sequence
 4 x_1, \dots, x_k . While proceeding along this sequence, the robot receives obser-
 5 vations y_1, \dots, y_k . Our goal is to estimate the state of each cell $m_i \in \{0, 1\}$
 6 (aka “the map” $m = (m_1, m_2, \dots,)$)

$$P(m | x_1, \dots, x_k, y_1, \dots, y_k) = \prod_i P(m_i | x_1, \dots, x_k, y_1, \dots, y_k). \quad (4.20)$$

- 7 This is called the “static state” Bayes filter and is conceptually exactly the
 8 same as the recursive application of Bayes rule in Chapter 2 for detecting
 9 whether the door was open or closed.

- 10 We will use a short form to keep the notation clear

$$y_{1:k} = (y_1, y_2, \dots, y_k);$$

- 11 the quantity $x_{1:k}$ is defined similarly. As usual we will use a recursive Bayes
 12 filter to compute this probability as follows.

$$\begin{aligned} P(m_i | x_{1:k}, y_{1:k}) &\stackrel{\text{Bayes rule}}{=} \frac{P(y_k | m_i, y_{1:k-1}, x_{1:k}) P(m_i | y_{1:k-1}, x_{1:k})}{P(y_k | y_{1:k-1}, x_{1:k})} \\ &\stackrel{\text{Markov}}{=} \frac{P(y_k | m_i, x_k) P(m_i | y_{1:k-1}, \cancel{x_{1:k-1}})}{P(y_k | y_{1:k-1}, x_{1:k})} \\ &\stackrel{\text{Bayes rule}}{=} \frac{P(m_i | y_k, x_k) P(y_k | x_k) P(m_i | y_{1:k-1}, x_{1:k-1})}{P(m_i | x_k) P(y_k | y_{1:k-1}, x_{1:k})} \\ &\stackrel{\text{Markov}}{=} \frac{P(m_i | y_k, x_k) P(y_k | x_k) P(m_i | y_{1:k-1}, x_{1:k-1})}{P(m_i) P(y_k | y_{1:k-1}, x_{1:k})}. \end{aligned} \quad (4.21)$$

- 13 We have a similar expression for the opposite probability

$$P(\neg m_i | x_{1:k}, y_{1:k}) = \frac{P(\neg m_i | y_k, x_k) P(y_k | x_k) P(\neg m_i | y_{1:k-1}, x_{1:k-1})}{P(\neg m_i) P(y_k | y_{1:k-1}, x_{1:k})}.$$

- 14 Let us take the ratio of the two to get

$$\begin{aligned} \frac{P(m_i | x_{1:k}, y_{1:k})}{P(\neg m_i | x_{1:k}, y_{1:k})} &= \frac{P(m_i | y_k, x_k)}{P(\neg m_i | y_k, x_k)} \frac{P(m_i | y_{1:k-1}, x_{1:k-1})}{P(\neg m_i | y_{1:k-1}, x_{1:k-1})} \frac{P(\neg m_i)}{P(m_i)} \\ &= \underbrace{\frac{P(m_i | y_k, x_k)}{1 - P(m_i | y_k, x_k)}}_{\text{uses observation } y_k} \underbrace{\frac{P(m_i | y_{1:k-1}, x_{1:k-1})}{1 - P(m_i | y_{1:k-1}, x_{1:k-1})}}_{\text{recursive term}} \underbrace{\frac{1 - P(m_i)}{P(m_i)}}_{\text{prior}}. \end{aligned} \quad (4.22)$$

- 15 This is called the odds ratio. Notice that the first term uses the latest obser-
 16 vation y_k , the second term can be updated recursively because it is a similar

1 expression as the left-hand side and the third term is a prior probability of the
 2 cell being occupied/non-occupied. Let us rewrite this formula using the log-
 3 odds-ratio that makes implementing it particularly easy. The log-odds-ratio of
 4 the probability $p(x)$ of a binary variable x is defined as

$$l(x) = \log \frac{p(x)}{1 - p(x)}, \text{ and } p(x) = 1 - \frac{1}{1 + e^{l(x)}}.$$

5 The product in (4.22) now turns into a sum as

$$l(m_i | y_{1:k}, x_{1:k}) = l(m_i | y_k, x_k) + l(m_i | y_{1:k-1}, x_{1:k-1}) - l(m_i). \quad (4.23)$$

6 This expression is used to update the occupancy of each cell. The term

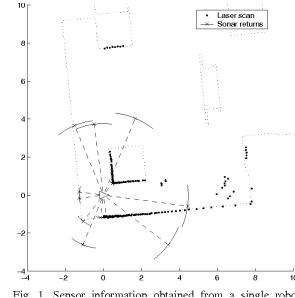
$$\text{sensor model} = l(m_i | y_k, x_k)$$

7 is different for different sensors and we will investigate it next.

8 4.3.2 Sensor models

9 **Sonar** This works by sending out an ultrasonic chirp and measuring the time
 10 between emission and reception of the signal. The time gives an estimate of
 11 the distance of an object to the robot.

12



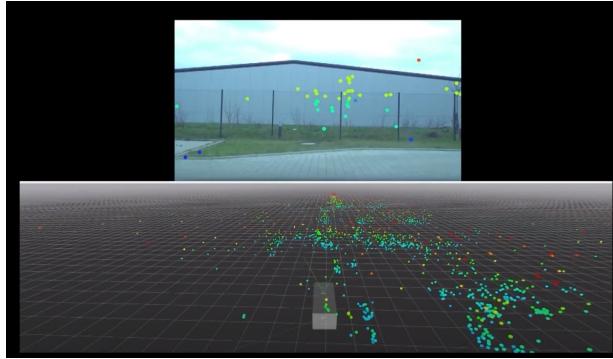
13 The figure above shows a typical sonar sensor (the two “eyes”) on a low-cost
 14 robot. Data from the sensor is shown on the right, a sonar is a very low
 15 resolution sensor and has a wide field of view, say 15 degrees, i.e., it cannot
 16 differentiate between objects that are within 15 degrees of each other and
 17 registers them as the same point. Sophisticated sonar technology is used today
 18 in marine environments (submarines, fish finders, detecting mines etc.).

19 **Radar** works in much the same way as a sonar except that it uses pulses
 20 of radio waves and measures the phase difference between the transmitted
 21 and the received signal. This is a very versatile sensor (it was invented by the
 22 US army to track planes and missiles during World War II) but is typically
 23 noisy and requires sophisticated processing to be used for mainstream robotics.
 24 Autonomous cars, collision warning systems on human-driven cars, weather
 25 sensing, and certainly the military use the radar today. The following picture
 26 and the video https://www.youtube.com/watch?v=hwKUcu_7F9E will give
 27 you an appreciation of the kind of data that a radar records. Radar is a very

?

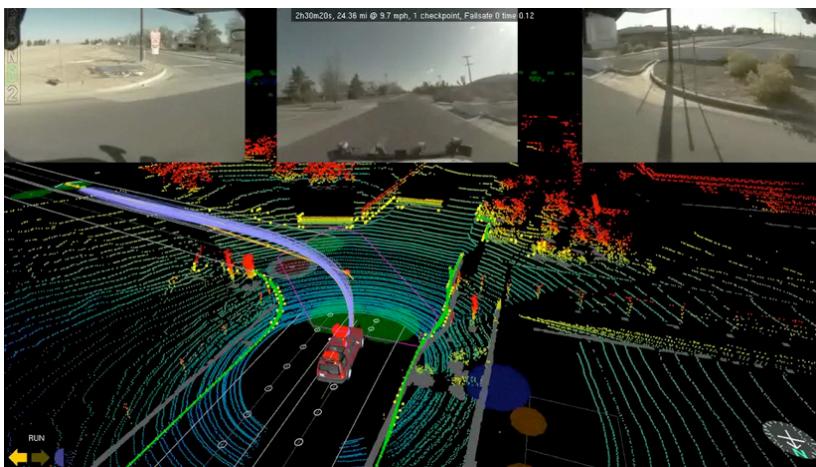
We assumed that the map was static. Can you think of why (4.23) automatically lets us handle some moving objects? Think of what the prior odds $l(m_i)$ does to the log-odds-ratio $l(m_i | y_{1:k}, x_{1:k})$.

- 1 long range sensor (typically 150 m) and works primarily to detect metallic
 2 objects.



3

4 **LiDAR** LiDAR, which is short for Light Detection and Ranging, (<https://en.wikipedia.org/wiki/Lidar>)
 5 is a portmanteau of light and radar. It is a sensor that uses a pulsed laser as the
 6 source of illumination and records the time it takes (nanoseconds typically) for
 7 the signal to return to the source. See <https://www.youtube.com/watch?v=NZKvf1cXe8s>
 8 for how the data from a typical LiDAR (Velodyne) looks like. While a Velodyne
 9 contains an intricate system of rotating mirrors and circuitry to measure
 10 time elapsed, there are new solid state LiDARs that are rapidly evolving to
 11 match the needs of the autonomous driving industry. Most LiDARs have a
 12 usable range of about 100 m.



13

14 **A typical autonomous car** This is a picture of MIT's entry named Talos to
 15 the DARPA Urban Challenge ([https://en.wikipedia.org/wiki/DARPA_Grand_Challenge_\(2007\)](https://en.wikipedia.org/wiki/DARPA_Grand_Challenge_(2007)))
 16 which was a competition where teams had to traverse a 60 mile urban route
 17 within 6 hours, while obeying traffic laws, understanding incoming vehicles
 18 etc. Successful demonstrations by multiple teams led by (CMU, Stanford,
 19 Odin College, MIT, Penn and Cornell) in this competition jump-started the
 20 wave of autonomous driving. While the number of sensors necessary to drive
 21 well has come down (Tesla famously does not like to use LiDARs and rely ex-
 22 clusively on cameras and radars), the type of sensors and challenges associated
 23 with them remain essentially the same.

❶ Waymo's autonomous car





1

2 4.3.3 Back to sensor modeling

3 Let us go back to understanding our sensor model $l(m_i | y_k, x_k)$ where m_i
 4 is a particular cell of the occupancy grid, y_k and x_k are the observations and
 5 robot position/orientation at time k .

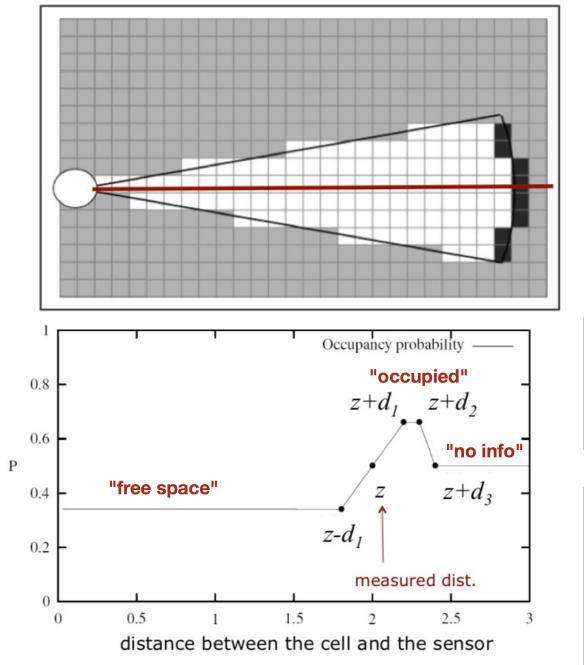
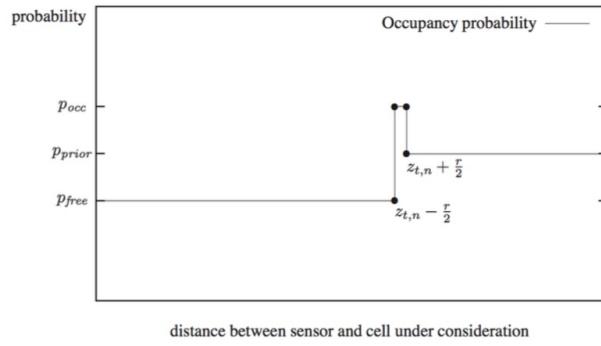


Figure 4.4: **Model for sonar data.** (Top) A sonar gives one real-valued reading corresponding to the distance measured along the red axis. (Bottom) if we travel along the optical axis, the occupancy probability $P(m_i | y_k = z, x_k)$ can be modeled as a spike around the measured value z . It is very important to remember that range sensors such as sonar gives us three kinds of information about this ray: (i) all parts of the environment up to $\approx z$ are *unoccupied* (otherwise we would not have recorded z), (ii) there is some object at z which resulted in the return, (iii) but we do not know anything about what is behind z . So incorporating a measurement y_k from a sonar/radar/lidar involves not just updating the cell which corresponds to the return, but also updating the occupancy probabilities of every grid cell along the axis.

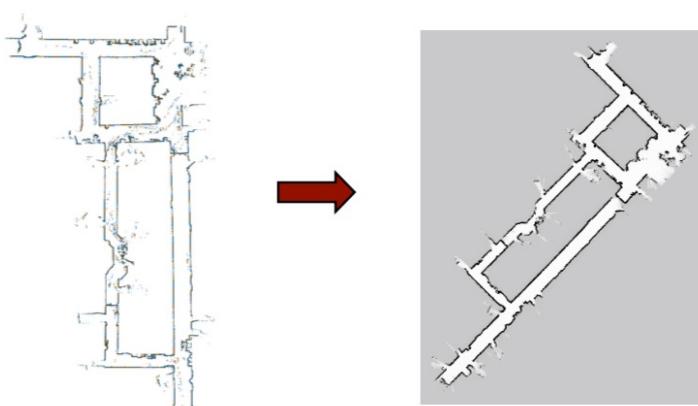


Figure 4.5: (Left) A typical occupancy grid created using a sonar sensor by updating the log-odds-ratio $l(m_i | x_{1:k}, y_{1:k})$ for all cells i for multiple time-steps k . At the end of the map building process, if $l(m_i | x_{1:k}, y_{1:k}) > 0$ for a particular cell, we set its occupancy to 1 and to zero otherwise, to get the maximum-likelihood estimate of the occupancy grid on the right.

- 1 **LiDAR model** When we say that a LiDAR is a more accurate sensor than
- 2 the sonar, what we really mean is that the sensor model $P(m_i | y_k, x_k)$ looks
- 3 as follows.

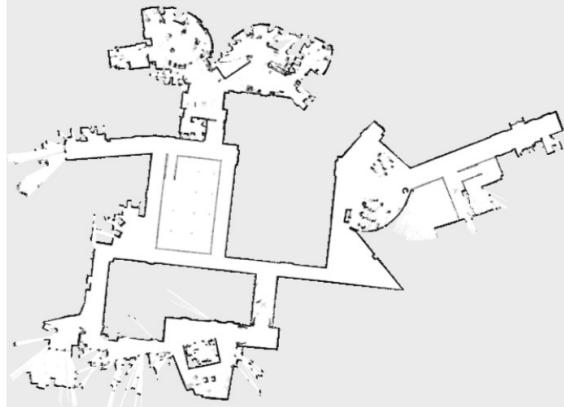


- 4
- 5 As a result, we can create high-resolution occupancy grids using a LiDAR.



- 6

Example: MIT CSAIL 3rd Floor

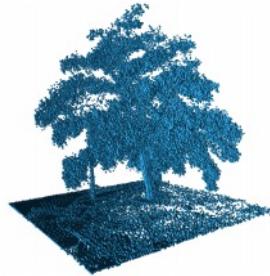


1

❸ How will you solve the localization problem given the map? In other words, if we know the occupancy grid of a building as estimated in a prior run, and we now want to find the position/orientation of the robot traveling in this building, how show we use these sensors?

2 4.4 3D occupancy grids

3 Two-dimensional occupancy grids are a fine representation for toy problems
 4 but they run into some obvious problems: since the occupancy grid is a “top view” of the world, we cannot represent non-trivial objects in it correctly
 5 (a large tree with a thin trunk eats up all the free space). We often desire a
 6 fundamentally three-dimensional representation of the physical world.
 7

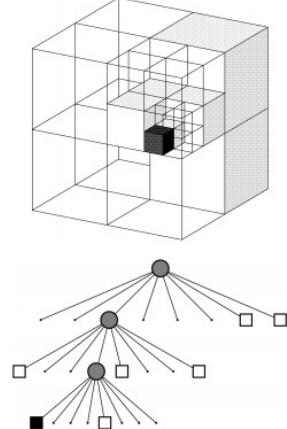


8

9 We could simply create cells in 3D space and our method for occupancy
 10 grid would work but this is no longer computationally cheap. For instance,
 11 if we want to build a map of Levine Hall (say 50 m × 50 m area and height
 12 of 25 m), a 3D grid map with a resolution of 5 cm × 5 cm × cm would have
 13 about 500 million cells (if we store a float in each cell this map will require
 14 about 2 GB memory). It would be cumbersome to carry around so many cells
 15 and update their probabilities after each sensor reading (a Velodyne gives data
 16 at about 30 Hz). More importantly, observe that most of the volume inside
 17 Levine is free space (inside of offices, inner courtyard etc.) so we do not really
 18 need fine resolution in those regions.

19 **Octrees** We would ideally have an occupancy grid whose resolution adapts
 20 with the kind of objects that are detected by the sensors. If nearby cells are
 21 empty we want to collapse them together to save on memory and computation,

1 on the other hand, if nearby cells are all occupied, we want to *refine* the
 2 resolution in that area so has to more accurately discern the shape of the
 3 underlying objects. Octrees are an efficient representation for 3D volumes.



4

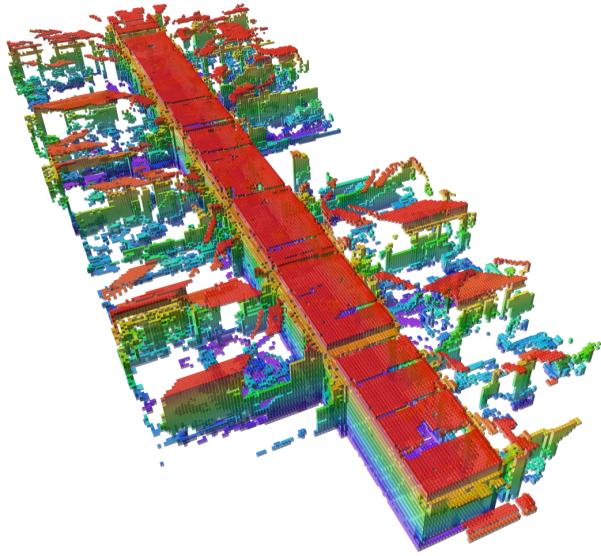
5 An octree is a hierarchical data structure that recursively sub-divides the 3D
 6 space into octants and allocates volumes as needed for a particular data point
 7 observed by a range sensor. It is analogous to a kd-tree. Imagine if the entire
 8 space in the above picture were empty (the tree only has a root node), and we
 9 receive a reading corresponding to the dark shaded region. An octree would
 10 sub-divide the space starting from the root (each node in the tree populates
 11 is the parent of its eight child octants) recursively until some pre-determined
 12 minimum resolution is reached. This leaf node is grid cell; notice how different
 13 cells in the octree have different resolutions. Occupancy probabilities of each
 14 leaf node are updated using the same formula as that of (4.23). A key point
 15 here is that octrees are designed for accurate sensors (LiDARs) where there is
 16 not much noise in the observations returned by the sensor (and thereby we do
 17 not refine unnecessary parts of the space)

18 Octrees are very efficient at storing large map, I expect you can store the
 19 entire campus of Penn in about a gigabyte. Ray tracing (following all the cells
 20 m_i in tree along the axis of the sensor in Figure 4.4) is harder in this case but
 21 there are efficient algorithms devised for this purpose. An example OctoMap
 22 (an occupancy map created using an Octree) of a building on the campus of
 23 the University of Freiburg is shown below.

❶ You can find LiDAR maps of the entire United States (taken from a plane) at <https://www.usgs.gov/core-science-systems/ngp/3dep>

24 4.5 Local Map

25 In this chapter, we primarily discussed occupancy grids of static environments
 26 as the robot moves around in the environment. The purpose of doing so is lo-
 27 calization, namely, finding the pose of the robot by comparing the observa-
 28 tions of the sensors with the map (think of the particle filter localization example
 29 in Chapter 3). In typical problems, we often maintain two kinds of maps, (i)
 30 a large occupancy grid for localization (say as big as city), and (ii) another
 31 smaller map, called the local map, that is used to maintain the locations of



¹ objects (typically objects that can move) in the vicinity of the robot, say a 100
² m × 100 m area.

³ The local map is used for planning and control purposes, e.g., to check if
⁴ the planned trajectory of the robot does not collide with any known obstacles.
⁵ See an example of the local map at the 1:42 min mark at
⁶ <https://www.youtube.com/watch?v=2va15BE-7lQ>. Some people also call the
⁷ local map a “cost map” because occupied cells in the local map indicate a
⁸ high collision cost of moving through that cell. The local map is typically
⁹ constructed in the body frame and evolves as the robot moves around (objects
¹⁰ appear in the front of the robot and are spawned in the local map and disappear
¹¹ from the map at the back as the robot moves forward).

You should think of the map (and especially the local map) as the filtering estimate of the locations of various objects in the vicinity of the robot computed on the basis of multiple observations received from the robot’s sensors.

¹² 4.6 Discussion

¹³ Occupancy grids are a very popular approach to represent the environment
¹⁴ given the poses of the robot as it travels in this environment. We can also
¹⁵ use occupancy grids to localize the robot in a future run (which is usu-
¹⁶ ally the purpose of creating them). Each cell in an occupancy grid stores
¹⁷ the posterior probability of the cell being occupied on the basis of mul-
¹⁸ tiple observations $\{y_1, \dots, y_k\}$ from respective poses $\{x_1, \dots, x_k\}$. This
¹⁹ is a very efficient representation of the 3D world around us with the one
²⁰ caveat that each cell is updated independently of the others. But since one



Figure 4.6: The output of perception modules for a typical autonomous vehicle (taken from <https://www.youtube.com/watch?v=tiwVMrTLUWg>. The global occupancy grid is shown in gray (see the sides of the road). The local map is not shown in this picture but you can imagine that it has occupied voxels at all places where there are vehicles (purple boxes) and other stationary objects such as traffic light, nearby buildings etc. Typically, if we know that so and so voxel corresponds to a vehicle, we run an Extended Kalman Filter for that particular vehicle to estimate the voxels in the local map that it is likely to be in, in the next time-instant. The local map is a highly dynamic data structure that is rich in information necessary for planning trajectories of the robot.

- ¹ gets a large amount of data from typical range sensors (a 64 beam Velodyne (<https://velodynelidar.com/products/hdl-64e>) returns about a 2 million points/sec and cheaper versions of this sensor will cost about \$100), this caveat
- ² does not hurt us much in practice. You can watch this talk
- ³ (https://www.youtube.com/watch?v=V8JMwE_L5s0) by the head of Uber's
- ⁴ autonomous driving group to get more perspective about localization and
- ⁵ mapping.