Unit V – ROBOTICS, ETHICS AND SAFETY IN AI: Robotics:Robots, Robot hardware Robotic perception, Alternative ,robotic frameworks,Application domains **.Ethics and Safety in AI**:Limits of AI,Ethics of AI: Surveillance,Security and privacy,Fairness and bias,Trust and transparency,AI safety.

"5.1 Robotics:

5.1.1 Robots

5.1.2 Robot hardware robotic perception

5.1.3 Alternative

5.1.4 Robotic frameworks

5.1.5 Application

In which agents are endowed with sensors and physical effectors with which to move about and make mischief in the real world

5.1.1 ROBOTS

- Robots are physical agents that perform tasks by manipulating the physical world.
- To do so, they are equipped with effectors such as legs, wheels, joints, and grippers.
- Effectors are designed to assert physical forces on the environment.
- Robots are also equipped with sensors, which enable them to perceive their environment.
- Present-day robotics employs a diverse set of sensors, including cameras, radars, lasers, and microphones to measure the state of the environment
- Robots operate in environments that are partially observable and stochastic: cameras cannot see around corners, and gears can slip.
- The people acting in that same environment are unpredictable, so the robot needs to make predictions about them.
- Robots usually model their environment with a continuous state space (the robot's position has continuous coordinates) and a continuous action space (the amount of current a robot sends to its motor is also measured in continuous units)
- Some robots operate in high dimensional spaces:
- cars need to know the position, orientation, velocity of themselves and the nearby agents;
- robot arms have six or seven joints that can each be independently moved; and robots that mimic the human body have hundreds of joints.
- Robotic learning is constrained because the real world stubbornly refuses to operate faster than real time.
- In a simulated environment, it is possible to use learning algorithms (such asthe Q-learning algorithm) to learn in a few hours from millions of trials.
- In a real environment, it might take years to run these trials, and the robot cannot risk (and thus cannot learn from) a trial that might cause harm.
- Transferring what has been learned in simulation to a real robot in the real world—the sim-to-real problem—is an active area of research
- Robotics brings together many of the concepts
    - **probabilistic state estimation, perception, planning, unsupervised learning, reinforcement learning, and game theory.**

- For some of these concepts robotics serves as a challenging example application.

5.1.2 ROBOT HARDWARE

- Types of robots from the hardware perspective
- Sensing the world
- Producing motion

✓ **Types of robots from the hardware perspective**

- Mobile robots are those that use wheels, legs, or rotors to move about the environment.
- Quadcopter drones are a type of unmanned aerial vehicle (UAV); autonomous underwater vehicles (AUVs) roam the oceans
- Legged robots are meant to traverse rough terrain that is inaccessible with wheels

(a) (b)

Figure 26.2 (a) NASA's Curiosity rover taking a selfie on Mars. Image courtesy of NASA. (b) A Skydio drone accompanying a family on a bike ride. Image courtesy of Skydio.

Other kinds of robots include

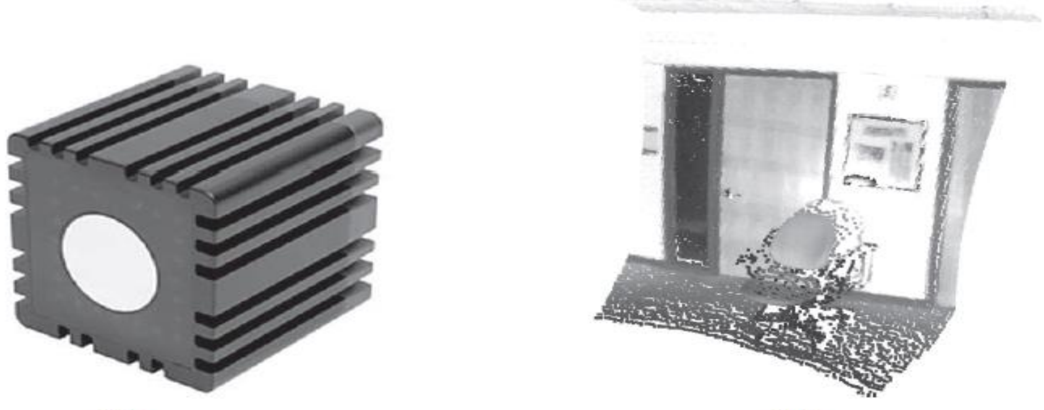
- Prostheses
- exoskeletons
- robots with wings, swarms
- intelligent environments in which the robot is the entire room

✓ **Sensing the World**

- Sensors are the perceptual interface between robot and environment
- Passive sensors, such as cameras, are true observers of the environment: they capture signals that are generated by other sources in the environment.
- Active sensors, such as sonar, send energy into the environment
- Range finders are sensors that measure the distance to nearby objects.
- Sonar sensors are active range finders that emit directional sound waves, which are reflected by objects, with some of the sound making it back to the sensor
- Stereo vision relies on multiple cameras to image the environment from slightly different viewpoints, analyzing the resulting parallax in these images to compute the range of surrounding objects.
- The Kinect is a popular low-cost sensor that combines a camera and a

- projector structured light, which projects a pattern of grid lines onto a scene.

- Figure 26.3(a) shows a time-of-flight camera.
- This camera acquires range images like the one shown in Figure 26.3(b) at up to 60 frames per second
- Autonomous cars often use scanning lidars (short for light detection and ranging)—active sensors that emit laser beams and sense the reflected beam, giving range measurements accurate to within a centimeter at a range of 100 meters



(a) (b)

Figure 26.3 (a) Time-of-flight camera; image courtesy of Mesa Imaging GmbH. (b) 3D range image obtained with this camera. The range image makes it possible to detect obstacles and objects in a robot's vicinity. Image courtesy of Willow Garage, LLC.

- Radar is often the range finding sensor of choice for air vehicles (autonomous or not).
- Radar sensors can measure distances up to kilometers
- On the close end of range sensing are tactile sensors such as whiskers, bump panels, and touch-sensitive skin
- A second important class is location sensors.
- Most location sensors use range sensing as a primary component to determine location.
- Outdoors, the Global Positioning System (GPS) is the most common solution to the localization problem
- At present, there are 31 operational GPS satellites in orbit, and 24 GLONASS satellites, the Russian counterpart
- Differential GPS involves a second ground receiver with known location, providing millimeter accuracy under ideal conditions.
- The third important class is proprioceptive sensors, which inform the robot of its own motion.
- To measure the exact configuration of a robotic joint, motors are often equipped
- with shaft decoders that accurately measure the angular motion of a shaft.
- On mobile robots, shaft decoders report wheel revolutions for odometry—the measurement of distance traveled.
- Inertial sensors, such as gyroscopes, reduce uncertainty by relying on the resistance of mass to the change of velocity.
- Other important aspects of robot state are measured by force sensors and torque sensors.
- These are indispensable when robots handle fragile objects or objects whose exact size and shape are unknown

✓ **Producing Motion**

- The mechanism that initiates the motion of an effector is called an actuator; examples :transmissions, gears, cables, and linkages.
- Common type of actuator is the electric actuator, which uses electricity to spin up a motor. These are predominantly used in systems that need rotational motion, like joints on a robot arm.
- Hydraulic actuators use pressurized hydraulic fluid (like oil or water) and pneumatic actuators use compressed air to generate mechanical motion.
- Actuators are often used to move joints, which connect rigid bodies (links).
- Arms and legs have such joints.
- In revolute joints, one link rotates with respect to the other. In prismatic
- joints, one link slides along the other.
- Both of these are single-axis joints (one axis of motion). Other kinds of joints include spherical, cylindrical, and planar joints, which are multi-axis joints.
- To interact with objects in the environment, robots use grippers.
- The most basic type of gripper is the parallel jaw gripper, with two fingers and a single actuator that moves the fingers together to grasp objects

5.1.3 Robotic Perception

- Localization and Mapping
- Other types of Perception
- Supervised and unsupervised learning in robot perception

- Perception is the process by which robots map sensor measurements into internal
- representations of the environment. Much of it uses the computer vision techniques
- Perception is difficult because sensors are noisy and the environment is partially observable, unpredictable, and often dynamic.
- As a rule of thumb, good internal representations for robots have three properties:
  1. They contain enough information for the robot to make good decisions.
  2. They are structured so that they can be updated efficiently.
  3. They are natural in the sense that internal variables correspond to natural state variables in the physical world.

- Kalman filters, HMMs, and dynamic Bayes nets can represent the transition and sensor models of a partially observable environment
- For robotics problems, we include the robot's own past actions as observed variables in the model

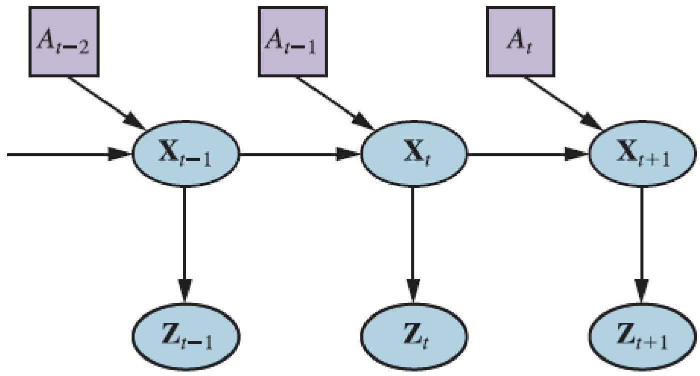


Figure 26.4 Robot perception can be viewed as temporal inference from sequences of actions and measurements, as illustrated by this dynamic decision network.

Notations Used:

- Xt - state of the environment at time t
- Zt - observation received at time
- At - action taken after the observation is received.

$$\mathbf{P}(\mathbf{X}_{t+1} | \mathbf{z}_{1:t+1}, a_{1:t})$$
$$= \alpha \mathbf{P}(\mathbf{z}_{t+1} | \mathbf{X}_{t+1}) \int \mathbf{P}(\mathbf{X}_{t+1} | \mathbf{x}_t, a_t)\ P(\mathbf{x}_t | \mathbf{z}_{1:t}, a_{1:t-1})\ d\mathbf{x}_t .$$

Sensor model

Transition model

This equation states that the posterior over the state variables $\mathbf{X}$ at time $t + 1$

Localization and Mapping

- Localization is the problem of finding out where things are—including the robot itself.
- To keep things simple, let us consider a mobile robot that moves slowly in a flat two dimensionalworld.
- Let us also assume the robot is given an exact map of the environment.

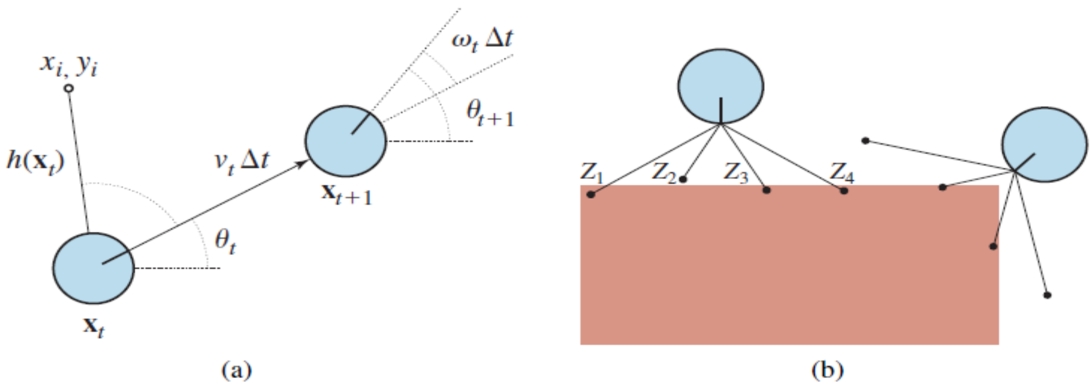


Figure 26.5 (a) A simplified kinematic model of a mobile robot. The robot is shown as a circle with an interior radius line marking the forward direction. The state $\mathbf{x}_t$ consists of the (x_t, y_t) position (shown implicitly) and the orientation θ_t. The new state $\mathbf{x}_{t+1}$ is obtained by an update in position of $v_t\Delta_t$ and in orientation of $\omega_t\Delta_t$. Also shown is a landmark at (x_i, y_i) observed at time t. (b) The range-scan sensor model. Two possible robot poses are shown for a given range scan (z_1, z_2, z_3, z_4). It is much more likely that the pose on the left generated the range scan than the pose on the right.

In the kinematic approximation, each action consists of the "instantaneous" specification of two velocities—a translational velocity v_t and a rotational velocity ω_t. For small time intervals Δt, a crude deterministic model of the motion of such robots is given by

$$\hat{\mathbf{X}}_{t+1+} = f(\mathbf{X}_{t,\underbrace{v_t,\omega_t}_{a_t}}) = \mathbf{X}_t + \begin{pmatrix} v_t \triangle t \cos\theta_t \\ v_t \triangle t \sin\theta_t \\ \omega_t \triangle t \end{pmatrix}.$$

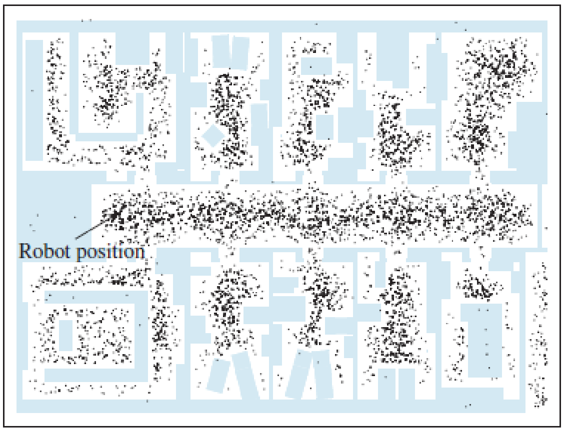
- we need a sensor model. We will consider two kinds of sensor models.
- The first assumes that the sensors detect stable, recognizable features of the environment called landmarks.
- To keep things simple, assume Gaussian noise with covariance giving us the sensor model

$$P(\mathbf{z}_t|\mathbf{x}_t) = N(\hat{\mathbf{z}}_t, \Sigma_z).$$

Localization using particle filtering is called Monte Carlo localization, or MCL. The MCL algorithm is an instance of the particle-filtering algorithm

function MONTE-CARLO-LOCALIZATIONa, z, N, $P(X'|X,\ v,\ \omega)$, $P(z|z^*)$, map
returns a set of samples, S, for the next time step
inputs: a, robot velocities v and ω
z, a vector of M range scan data points
$P(X'|X,\ v,\ \omega)$, motion model
$P(z|z^*)$, a range sensor noise model
map, a 2D map of the environment
persistent: S, a vector of N samples
local variables: W, a vector of N weights
S', a temporary vector of N samples

if S is empty **then**
  for $i = 1$ to N **do** // *initialization phase*
    $S[i] \leftarrow$ sample from $P(X_0)$
for $i = 1$ to N **do** // *update cycle*
  $S'[i] \leftarrow$ sample from $P(X'|X = S[i], v, \omega)$
  $W[i] \leftarrow 1$
  for $j = 1$ to M **do**
    $z^* \leftarrow$ RAYCAST(j, $X = S'[i]$, map)
    $W[i] \leftarrow W[i] \cdot P(z_j|\ z^*)$
  $S \leftarrow$ WEIGHTED-SAMPLE-WITH-REPLACEMENT(N, S', W)
return S



(a)

(b)

Monte Carlo localization, a particle filtering algorithm for mobile robot localization

Initial, global uncertainty.

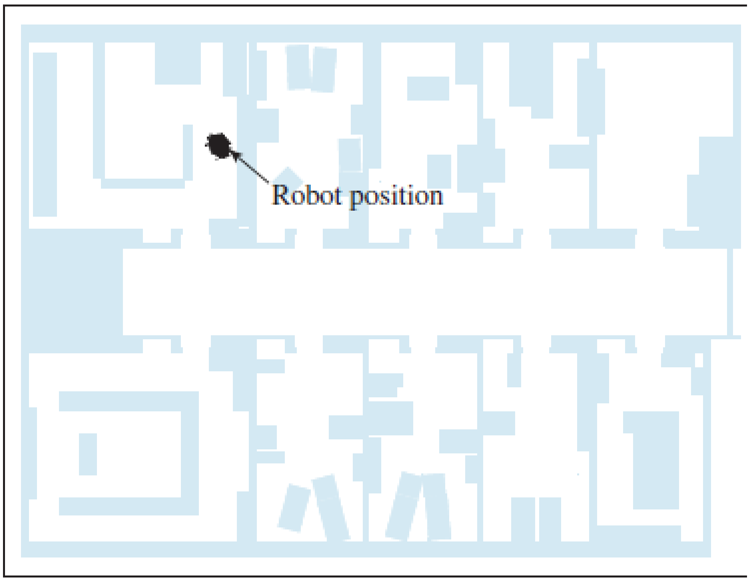
Bimodal uncertainty after navigating in corridor

The particles are uniformly distributed based on the prior, indicating global

uncertainty about the robot’s position

The first set of measurements

arrives and the particles form clusters in the areas of high posterior belief



(c)

Monte Carlo localization, a particle filtering algorithm for mobile robot localization

Unimodal uncertainty after entering a room and finding it to be distinctive

Enough measurements are available to push all the particles to a single location

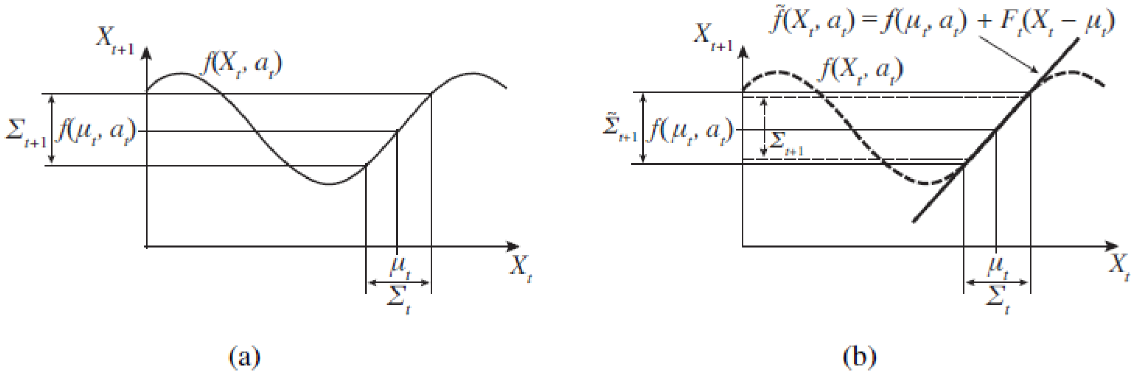


Figure 26.8 One-dimensional illustration of a linearized motion model: (a) The function f, and the projection of a mean μ_t and a covariance interval (based on Σ_t) into time $t+1$. (b) The linearized version is the tangent of f at μ_t. The projection of the mean μ_t is correct. However, the projected covariance $\tilde{\Sigma}_{t+1}$ differs from Σ_{t+1}.

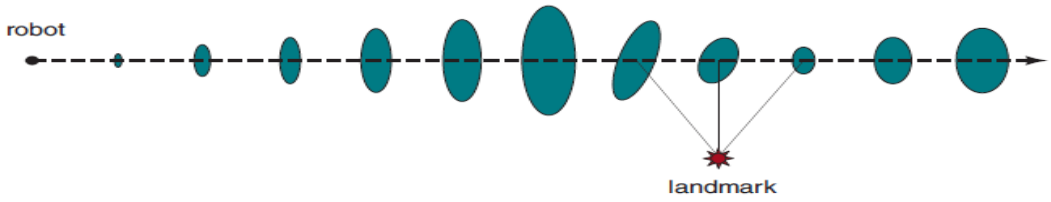


Figure 26.9 Localization using the extended Kalman filter. The robot moves on a straight line. As it progresses, its uncertainty in its location estimate increases, as illustrated by the error ellipses. When it observes a landmark with known position, the uncertainty is reduced.

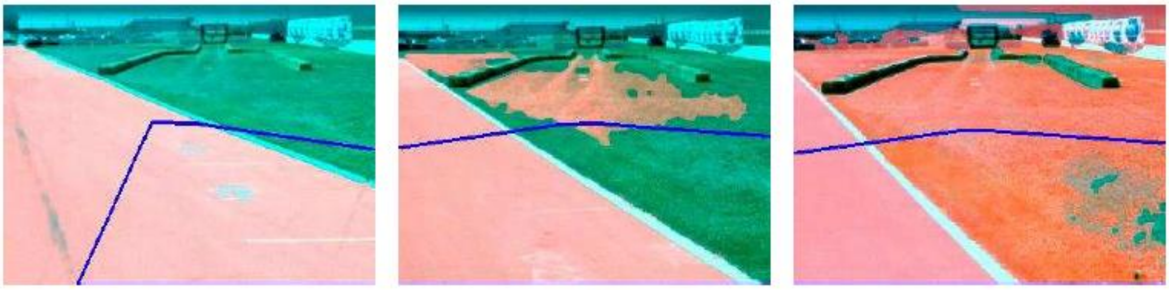
- ❖ As the robot moves, the uncertainty in its location estimate increases, as shown by the error ellipses.
- ❖ Its error decreases as it senses the range and bearing to a landmark with known location and increases again as the robot loses sight of the landmark
- ❖ The problem of needing to know the identity of landmarks is an instance of the data association problem
- ❖ In some situations, no map of the environment is available. Then the robot will have to acquire a map
- ❖ This problem is important for many robot applications, and it has been studied extensively under the name simultaneous localization and mapping, abbreviated as SLAM.

Other types of perception

- ❑ Not all of robot perception is about localization or mapping.
- ❑ Robots also perceive temperature, odors, sound, and so on.
- ❑ Many of these quantities can be estimated using variants of dynamic Bayes networks
- ❑ All that is required for such estimators are conditional probability distributions that characterize the evolution of state variables over time, and sensor models that describe the relation of measurements to state variables.
- ❑ Models required for such estimators are conditional probability distributions that characterize the evolution of state variables over time, and sensor models that describe the relation of measurements to state variables.
- ❑ It is also possible to program a robot as a reactive agent, without explicitly reasoning about probability distributions over states
- ❑ Probabilistic techniques outperform other approaches in many hard perceptual problems such as localization and mapping

Supervised and unsupervised Learning in Robot Perception

- ❑ Machine learning plays an important role in robot perception.
- ❑ If internal representation is not known then the common approach is to map high dimensional sensor streams into lower-dimensional spaces using unsupervised machine learning methods
- ❑ Such an approach is called low-dimensional embedding.
- ❑ Machine learning makes it possible to learn sensor and motion models from data, while simultaneously discovering a suitable internal representation



(a) (b) (c)

Figure 26.10 Sequence of "drivable surface" classifications using adaptive vision. (a) Only the road is classified as drivable (pink area). The V-shaped blue line shows where the vehicle is heading. (b) The vehicle is commanded to drive off the road, and the classifier is beginning to classify some of the grass as drivable. (c) The vehicle has updated its model of drivable surfaces to correspond to grass as well as road. Courtesy of Sebastian Thrun.

5.1.3 Alternative Robotic Frameworks

- Reactive Controllers
- Subsumption Architectures

View of robotics based on the notion of defining or learning a reward function, and having the robot optimize that reward function

This is a deliberative view of robotics, to be contrasted with a reactive view.

Reactive Controllers

- It is easier to set up a good policy for a robot than to model the world and plan. Then, instead of a rational agent, we have a reflex agent picture a legged robot that attempts to lift a leg over an obstacle.
- Give this robot a rule that says lift the leg a small height and move it forward, and if the leg encounters an obstacle, move it back and start again at a higher height

5.1.4 Application Domains

Robotic technology is already permeating our world, and has the potential to improve our independence, health, and productivity.

Example Applications

- HOME CARE
- HEALTH CARE
- SERVICES
- AUTONOMOUS CARS
- ENTERTAINMENT
- EXPLORATION AND HAZARDOUS ENVIRONMENTS
- INDUSTRY

✓ **HOME CARE**

- ❑ Robots have started to enter the home to care for older adults and people with motor impairments, assisting them with activities of daily living and enabling them to live more independently.
- ❑ These include wheelchairs and wheelchair-mounted arms like the Kinova arm
- ❑ Robots operated by brain–machine interfaces, which have been shown to enable people with quadriplegia to use a robot arm to grasp objects and even feed themselves
- ❑ Personal robots are meant to assist us with daily tasks like cleaning and organizing, freeing up our time

(a) (b)

Figure 26.33 (a) A patient with a brain–machine interface controlling a robot arm to grab a drink. Image courtesy of Brown University. (b) Roomba, the robot vacuum cleaner. Photo by HANDOUT/KRT/Newscom.

✓ **SERVICES**

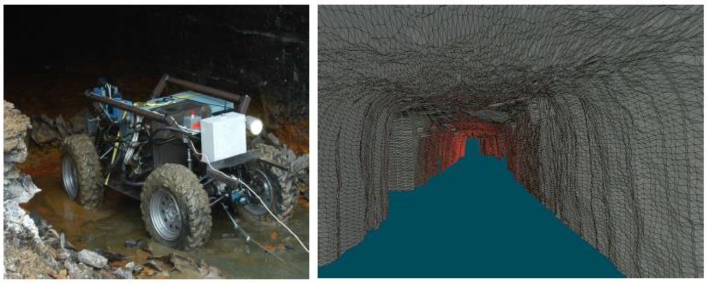
- ❑ Mobile robots help out in office buildings, hotels, and hospitals.
- ❑ Savioke has put robots in hotels delivering products like towels or toothpaste to your room.
- ❑ The Helpmate and TUG robots carry food and medicine in hospitals
- ❑ Diligent Robotics' Moxi robot helps out nurses with back-end logistical responsibilities.
- ❑ Co-Bot roams the halls of Carnegie Mellon University, ready to guide you to someone's office.
- ❑ We can also use telepresence robots like the Beam to attend meetings and conferences remotely, or check in on our grandparents.
- ❑ AUTONOMOUS CAR
- ❑ Some of us are occasionally distracted while driving, by cell phone calls, texts, or other distractions
- ❑ More than a million people die every year in traffic accidents.
- ❑ Further, many of us spend a lot of time driving and would like to recapture some of that time.
- ❑ All this has led to a massive ongoing effort to deploy autonomous cars.
- ❑ Prototypes have existed since the 1980s, but progress was stimulated by the 2005 DARPA Grand Challenge, an autonomous vehicle race over 200 challenging kilometers of unrehearsed desert terrain
- ❑ Figure 26.35(a) depicts BOSS, which in 2007 won the DARPA Urban Challenge

(a) (b)

Figure 26.35 (a) Autonomous car BOSS which won the DARPA Urban Challenge. Photo by Tangi Quemener/AFP/Getty Images/Newscom. Courtesy of Sebastian Thrun. (b) Aerial view showing the perception and predictions of the Waymo autonomous car (white vehicle with green track). Other vehicles (blue boxes) and pedestrians (orange boxes) are shown with anticipated trajectories. Road/sidewalk boundaries are in yellow. Photo courtesy of Waymo.

✓ **ENTERTAINMENT**

- Disney has been using robots (under the name animatronics) in their
- parks since 1963.
- Since 2009 a version called autonomatronics can generate autonomous actions. Robots also take the form of intelligent toys for children;
- quadrotors like Skydio's R1 from Figure 26.2(b) act as personal photographers and videographers, following us around to take action shots as we ski or bike.
- EXPLORATION AND HAZARDOUS ENVIRONMENTS
- Robots have gone where no human has gone before, including the surface of Mars. Robotic arms assist astronauts in deploying and retrieving satellites and in building the International Space Station.
- Robots also help explore under the sea.
- Figure 26.36 shows a robot mapping an abandoned coal mine, along with a 3D model of the mine acquired using range sensors.
- Robots have assisted people in cleaning up nuclear waste, most notably in Three Mile Island, Chernobyl, and Fukushima.
- Robots were present after the collapse of the World Trade Center, where they entered structures deemed too dangerous for human search and rescue crews."



(a) (b)

Figure 26.36 (a) A robot mapping an abandoned coal mine. (b) A 3D map of the mine acquired by the robot. Courtesy of Sebastian Thrun.

-

✓ **INDUSTRY**

- The majority of robots today are deployed in factories, automating tasks that
- are difficult, dangerous, or dull for humans.
- The majority of factory robots are in automobile factories.
- Automating these tasks is a positive in terms of efficiently producing what society needs.
- At the same time, it also means displacing some human workers from
- their jobs.
- This has important policy and economics implications—the need for retraining and education, the need for a fair division of resources

5.2.1 Limits of AI

- **The argument from informality**
- **The argument from disability**
- **The mathematical objection**
- **Measuring AI**

- **weak AI**—the idea that machines could act as if they were intelligent
- **strong AI**—the assertion that machines that do so are actually consciously thinking
- Alan Turing (1950), the first person to define AI, was also the first to raise possible objections to AI, foreseeing almost all the ones subsequently raised by others.

The argument from informality

- Turing's "argument from informality of behavior" says that human behavior is far too complex to be captured by any formal set of rules
- Philosopher Kenneth Sayre (1993) said "Artificial intelligence pursued within the cult of computationalism stands not even a ghost of a chance of producing durable results."
- The technology they criticize came to be called Good Old-Fashioned AI (GOFAI).
- GOFAI corresponds to the simplest logical agent design
- The embodied cognition approach claims that it makes no sense to consider the brain separately: cognition takes place within a body, which is embedded in an environment

The argument from disability

- The "argument from disability" makes the claim that "a machine can never do ."
- Computer chess expert David Levy predicts that by 2050 people will routinely fall in love with humanoid robots.
- Computers have done things that are "really new," making significant discoveries in astronomy, mathematics, chemistry, mineralogy, biology, computer science, and other fields, and creating new forms of art through style transfer (Gatys et al., 2016).

The mathematical objection

- Turing (1936) and Gödel (1931) proved that certain mathematical questions are in principle unanswerable by particular formal systems. Gödel's incompleteness theorem) is the most famous example of this
- Briefly, for any formal axiomatic framework powerful enough to do arithmetic, it is possible to construct a so-called Gödel sentence with the following properties:
- G(F)is a sentence of F, but cannot be proved within F.
- If F is consistent, then G(F) is true.

- Three of the problems with Lucas's claim. (Lucas assumes that humans can "change their minds" while computers cannot, but that is also false)
- First, an agent should not be ashamed that it cannot establish the truth of some sentence while other agents can.
- Second, Gödel's incompleteness theorem and related results apply to mathematics, not to computers.
- Third, Gödel's incompleteness theorem technically applies only to formal systems that are powerful enough to do arithmetic.

Measuring AI

- Alan Turing, in his famous paper "Computing Machinery and Intelligence" (1950), suggested that instead of asking whether machines can think, we should ask whether machines can pass a behavioral test, which has come to be called the Turing test
- The test requires a program to have a conversation (via typed messages) with an interrogator for five minutes.
- The interrogator then has to guess if the conversation is with a program or a person; the program passes the test if it fools the interrogator 30% of the time.
- Turing test competitions have led to better chatbots, but have not been a focus of research within the AI community
- Instead, AI researchers who crave competition are more likely to concentrate on playing chess or Go or StarCraft II, or taking an 8th grade science exam, or
- identifying objects in images. In many of these competitions, programs have reached or surpassed human-level performance, but that doesn't mean the programs are human-like
- The point is to improve basic science and technology and to provide useful tools, not to fool judges.

******