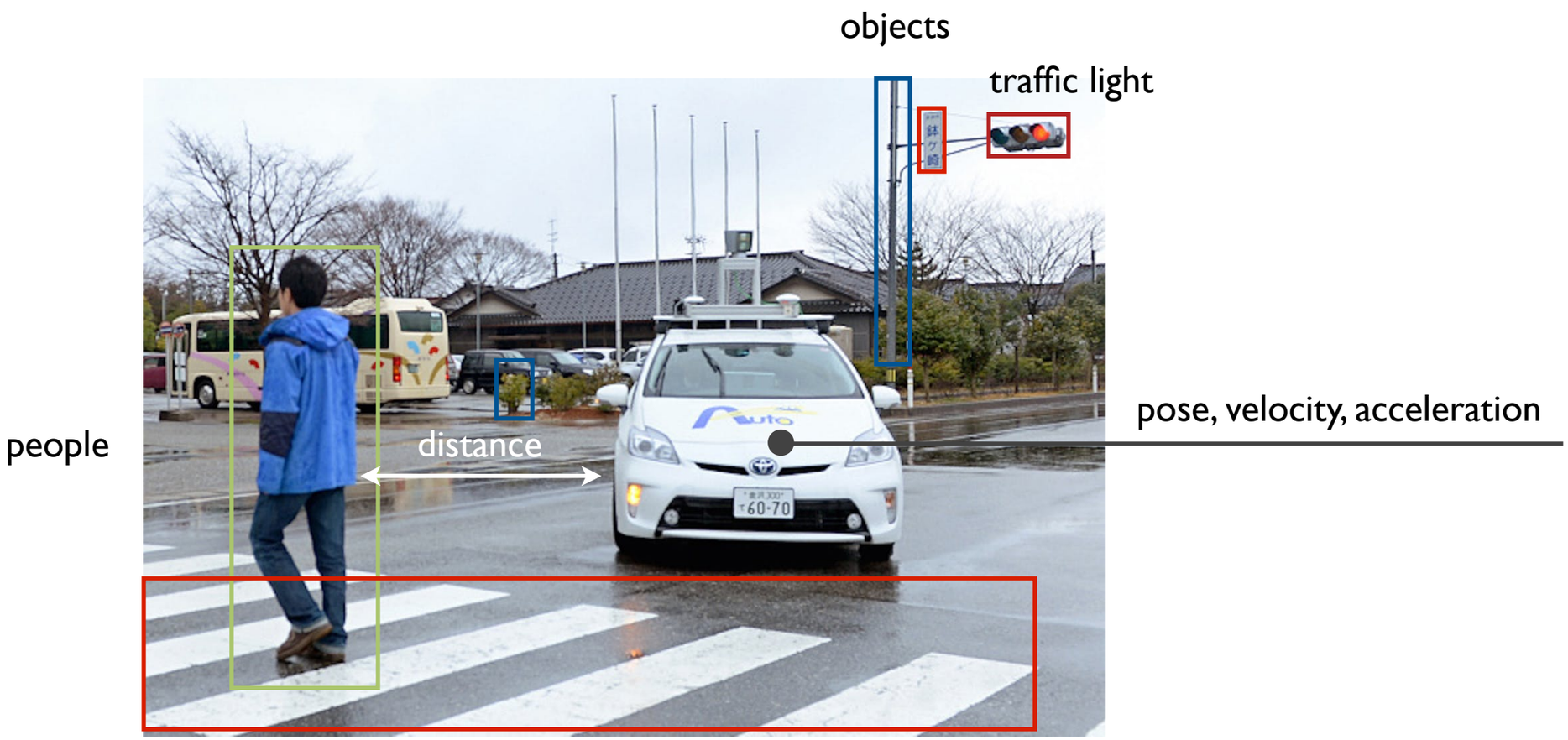
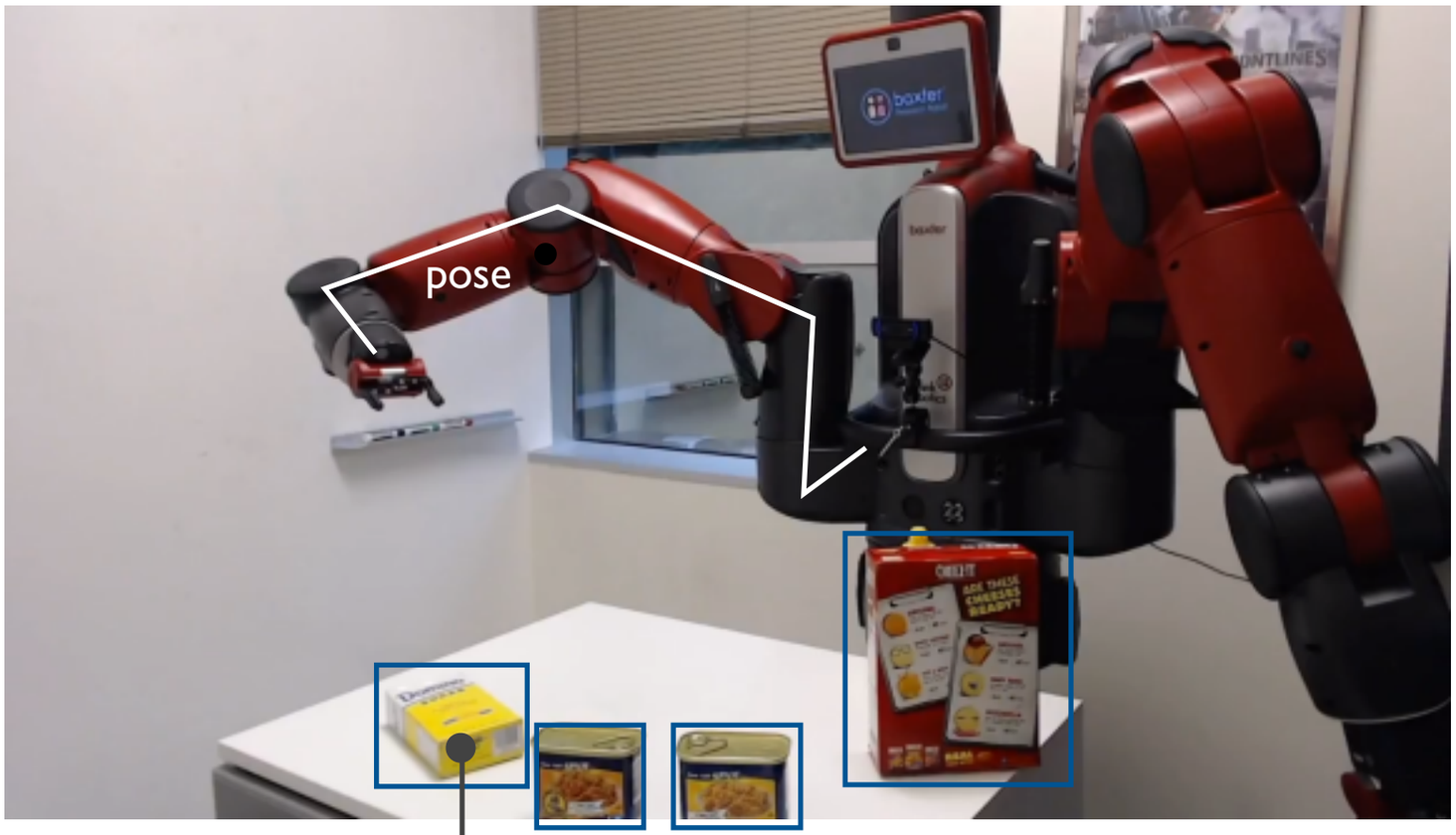
Sensors

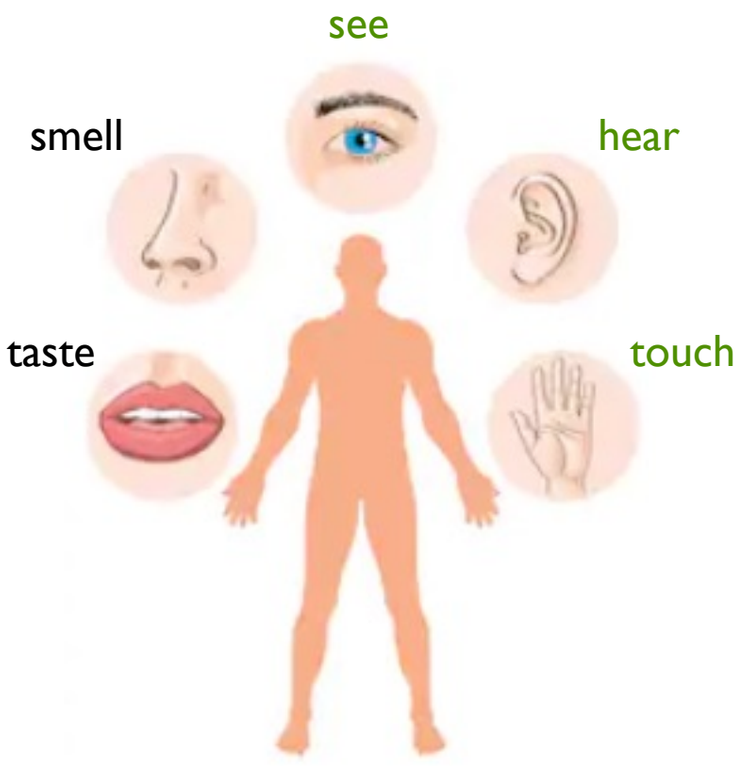
What does the robot sense?



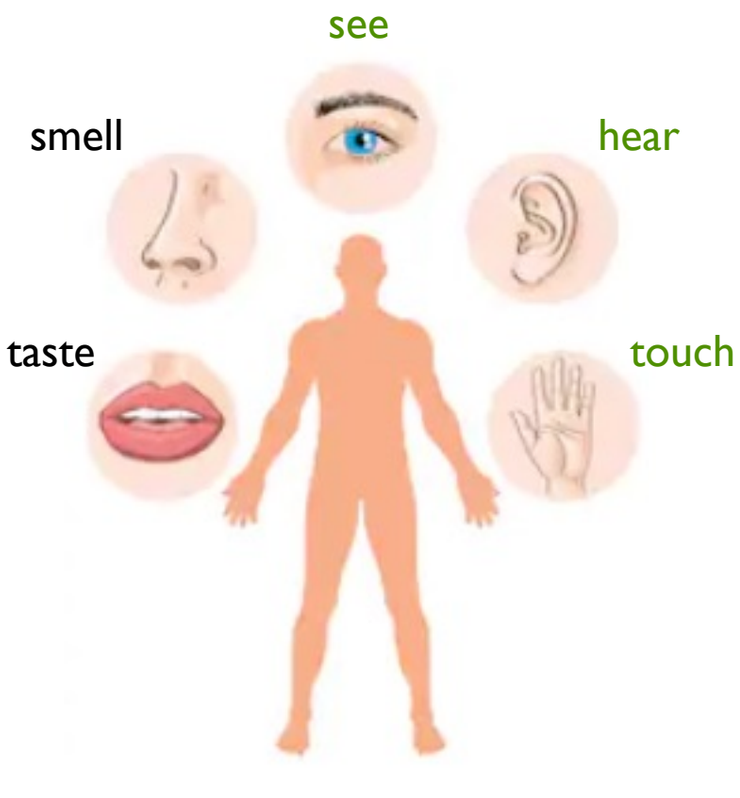
What does the robot sense?

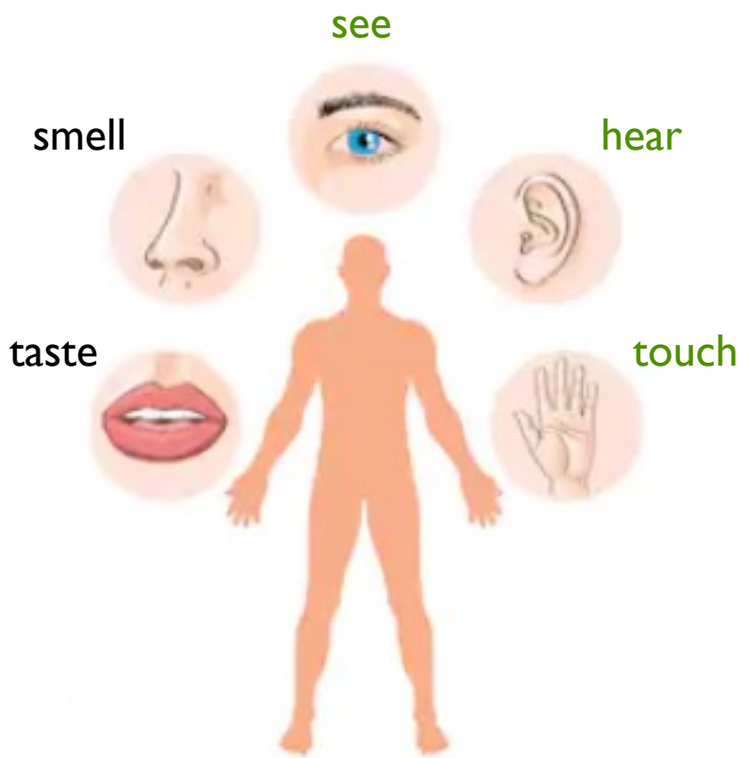


What does the human do?



What does the robot do?


see
smell
hear
taste
touch


see
smell
hear
taste
touch

All sensors are noisy! Some are less noisy than others.

Sensor performance.

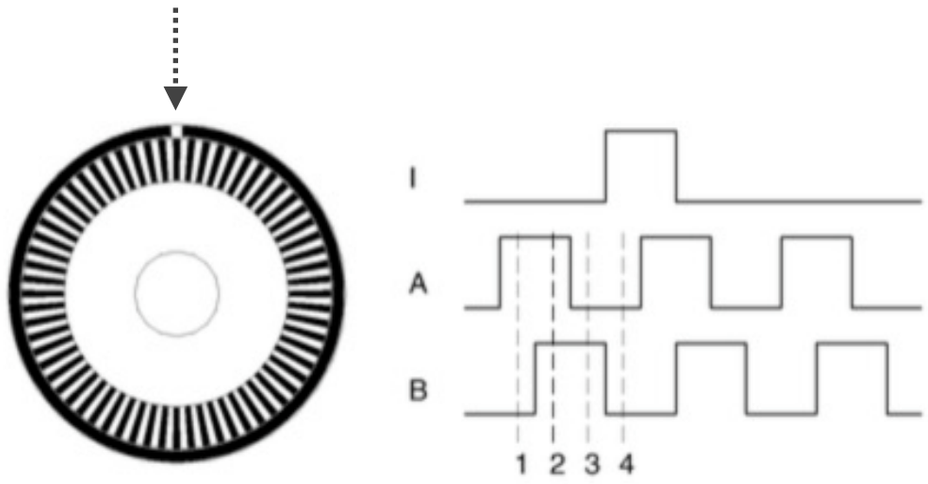
- Dynamic range. The dynamic range is the spread between the sensor's maximum and minimum values.
- Resolution. The resolution is the minimum difference between two values that the sensor can differentiate.
- Error. The error measures the absolute or the relative difference between the sensor's output measurement and the ground-truth value.
- Bandwidth or frequency. The frequency measures the speed at which the sensor can provide a sequence of stable measurements.
- Weight.
- Physical dimensions.
- Cost.

These performance metrics are usually reported in technical specifications. However, environmental factors may substantially influence the actual performance.

The camera has a minimum and a maximum operating distance. The imaging sensor has a resolution. The error is usually not explicitly specified, as visual perception is quite tolerate of minor errors.

The compass takes a few seconds to settle to a stable state. The frequency/bandwidth is low. It is slow.

Encoder. Optimal encoders measure angular position and speed. It is very accurate. Usually the measurement error is negligible.

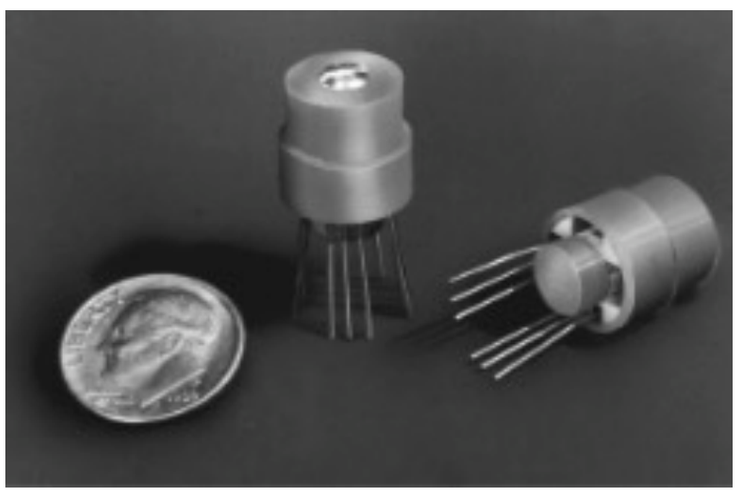


We use the encoder to get accurate measurement robot arm joint angles.

Odometry uses the encoder to measure the distance that the robot traverses by counting the number of wheel turns. The encoder counts the number of turns quite accurately. However, the distance measurement is nevertheless noisy, because wheels may slip. The relationship between wheel turns and distance traveled is noisy. In other words, encoders are accurate, but odometry is noisy.

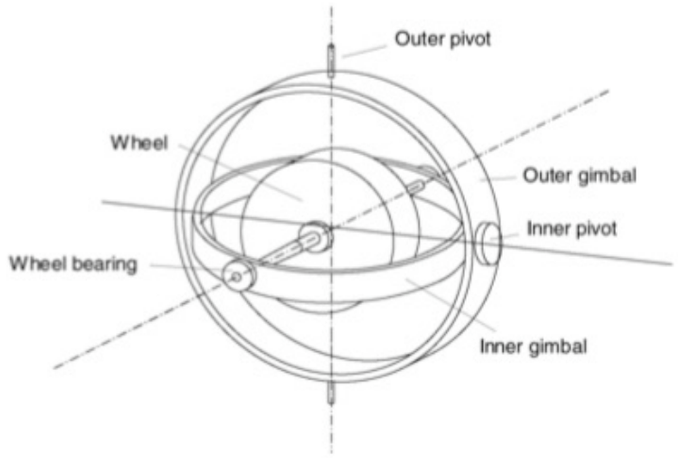
Compasses. Compasses measure the direction with respect to a magnetic field, e.g., that of the earth. Compass readings are quite noisy and somewhat slow, i.e., low-frequency. Further, it easily suffers from disturbance of the magnetic field by other magnetic objects and man-made structures.

compass



digital compass

Gyroscopes. Gyroscopes measure the 3-D orientation with respect to a fixed coordinate frame. Gyroscope is very accurate.



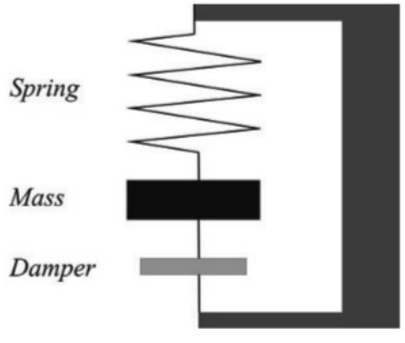
2-axis mechanical gyroscope

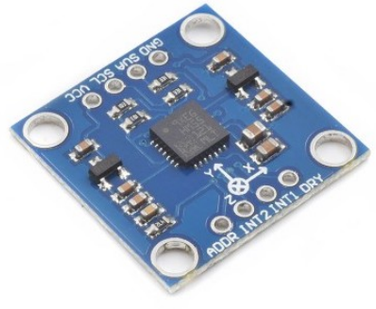
Accurate mechanical gyroscopes are very expensive navigational equipment for traveling long distances in the sea, in the air, or in the space. Optical gyroscopes are much cheaper and accessible.

Rate gyroscopes measure the angular velocity instead of the absolute orientation.

Accelerometers. Accelerometers measure the external forces acting a body: $F = kx = ma$. So

$$a = \frac{k}{m}x$$

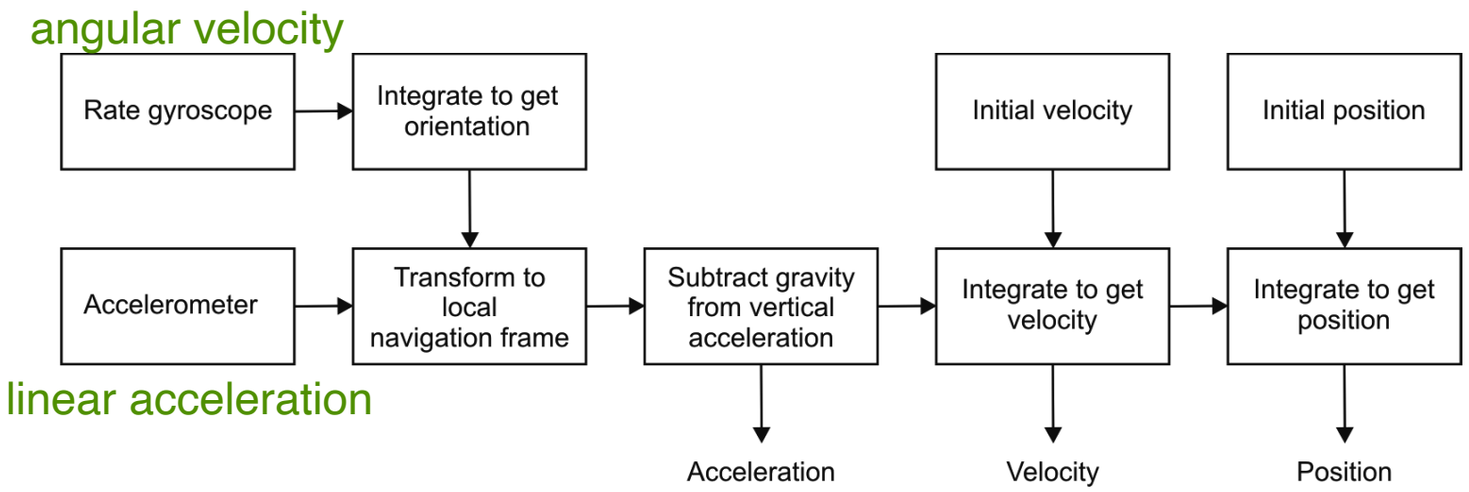




digital accelerometer

Accelerometers come in a wide range of accuracy and frequency, at different cost.

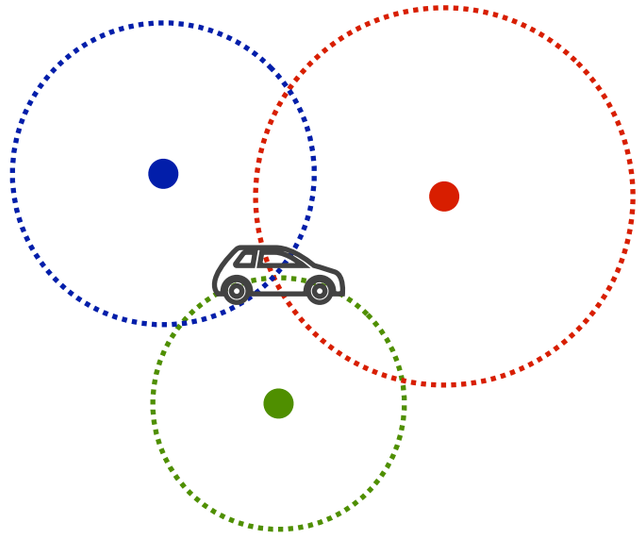
Inertial measurement units (IMUs). An IMU combines an accelerometer and a gyroscope to measure the relative 3-D pose, velocity, and acceleration of a moving body.



IMU measurements are relative. After some time of operation, errors accumulate, and measurements drift. They must be corrected through measurement with respect to an absolute frame, e.g., the GPS.

Global positioning system (GPS). The GPS measures the location on earth in an absolute frame. It relies on a system of 24 satellites. It requires distance measurement with respect to at least 4 satellites in order calculate the location through triangulation.

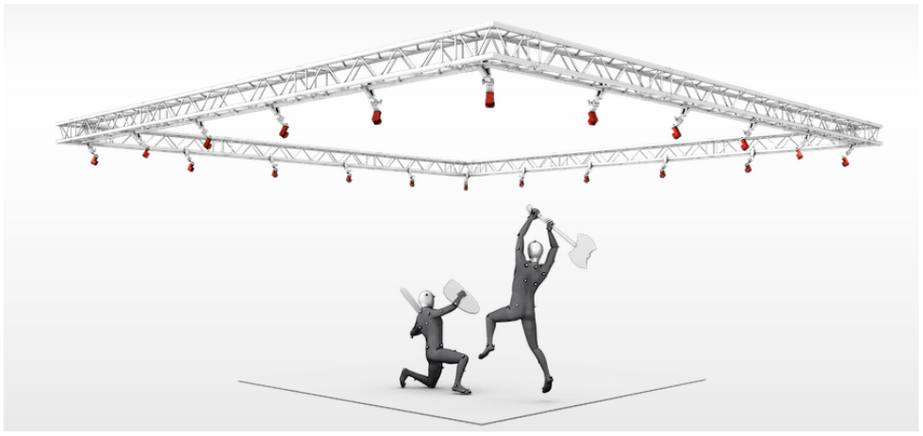
The diagram shows that to fix a location in the 2-D space, we need 3 distance measurements. By inference, to fix a location in the 3-D space, we need 4 distance measurements.



The standard GPS provides accuracy of approximately 10 m and is very fast. The differential GPS uses a static receiver at an exactly known location to correct the errors. The accuracy then improves to less than 1 m. Through more advanced error correction techniques or multiple static receivers, the accuracy can improve to 1 cm or less.

At least 4 satellites must be visible for GPS localization. This is an inherent limitation, for example, in indoor environments.

Of course, we can set up beacons indoors in lieu of satellites. Motion capture systems work exactly this way.



So far, we have considered sensing the robot's own pose, velocity, … Equally important is to sense the relationship between the robot and the environment, e.g., the distance.

Active ranging. A ranging sensor measures the distance. An active ranging sensor does so by emitting a beam of mechanic or electromagnetic wave into the environment and measures the response.

Two common measurement principles behind ranging sensors are time-of-flight and triangulation, but there are others.

Time of flight. The distance is proportional to the amount of time taken for the beam to reach the nearest object and reflect back to the receiver:

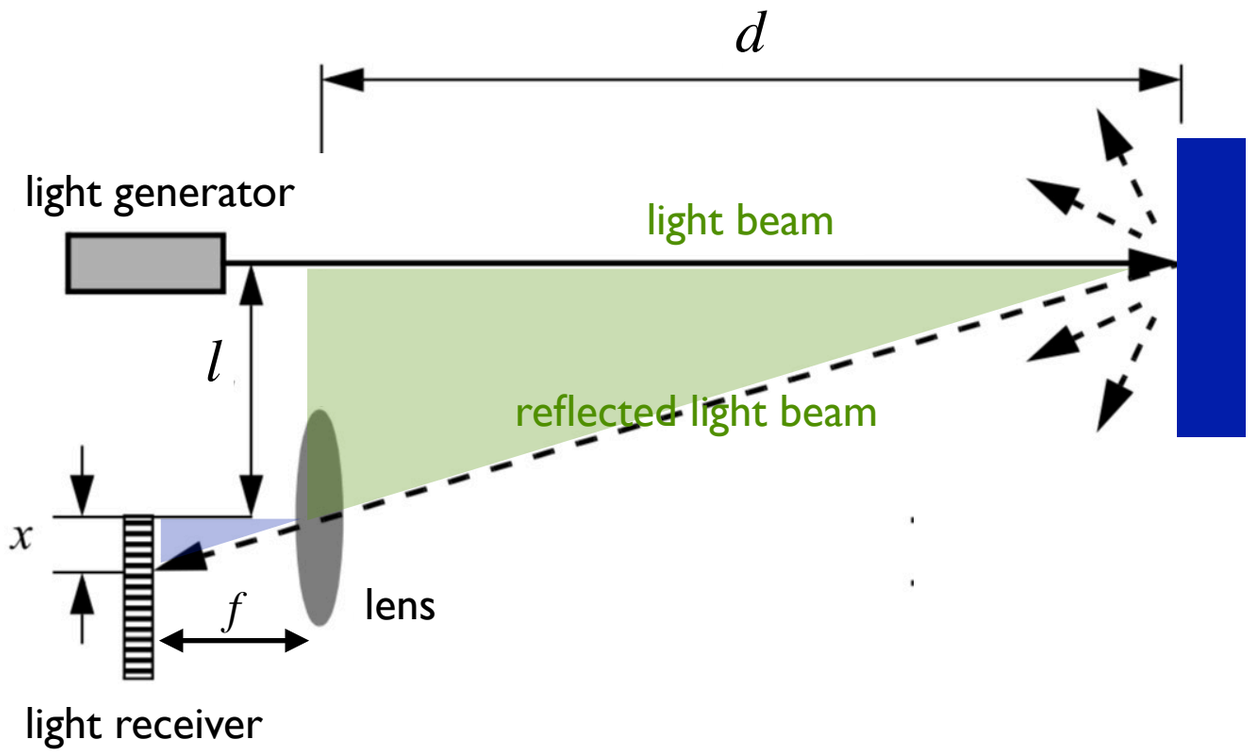
$$d = c \times t/2$$

speed of wave propagation

time of flight

The idea is simple. However, wave propagation may depend on the physical nature of the wave and many environmental factors, e.g., the medium of transport, temperature, … Further, some waves, e.g., light, travel at high speed. Accurately measuring the time of flight requires sensitive and potentially expensive mechanisms, especially when the **distance is small**.

Triangulation. Triangulation exploits the geometry. It projects a known wave pattern and locates the pattern on the receiver to determine the distance.



From the geometry of optics, we can work out

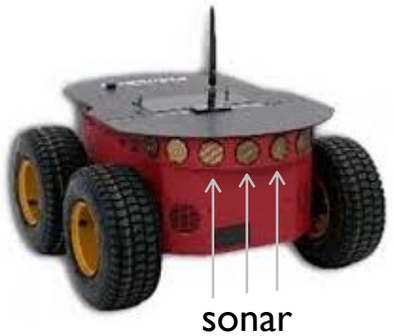
$$\frac{d}{l} = \frac{f}{x} \quad \text{and} \quad d = l\frac{f}{x}.$$

So since the measure measurement x is proportional to $1/d$, the measurement is more accurate, when d **is small**, i.e., the object is close, in contrast with the time-of-flight measurement.

Here, "more accurate" means higher resolution.

There are many types of waves: sound, radio, light, … They different in wavelength, frequency, absorption rate in the medium of travel, reflection properties upon hitting a surface,… These physical properties affect the performance of the sensor as well as physical dimensions, weight, and cost.

sound
(mechanical wave)



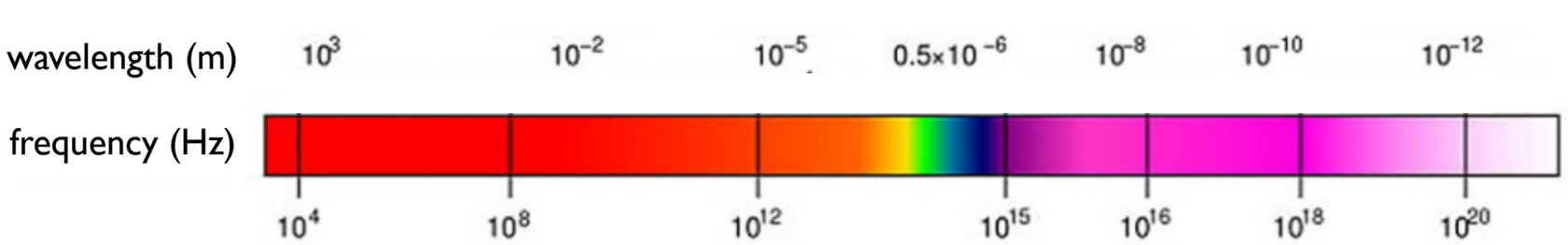
sonar

ultrasound distance sensor

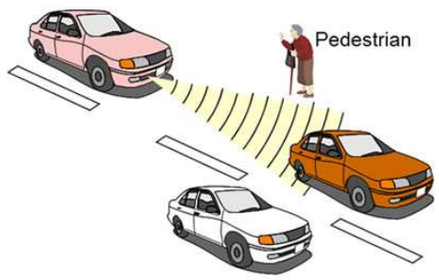
ultrasound image

electromagnetic wave

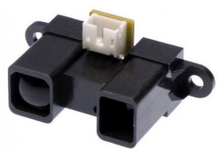
radio microwave infrared **visible** ultraviolet X-ray



RADAR

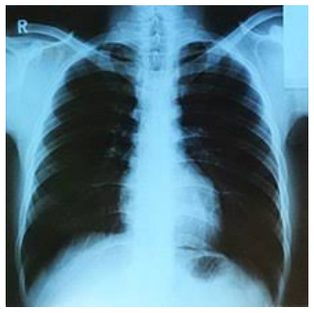


millimeter RADAR
(30-300 GHz)



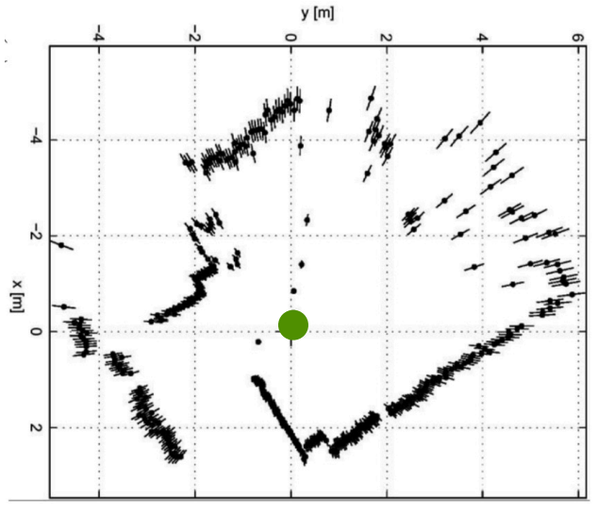
IR distance sensor
(~850 nm)

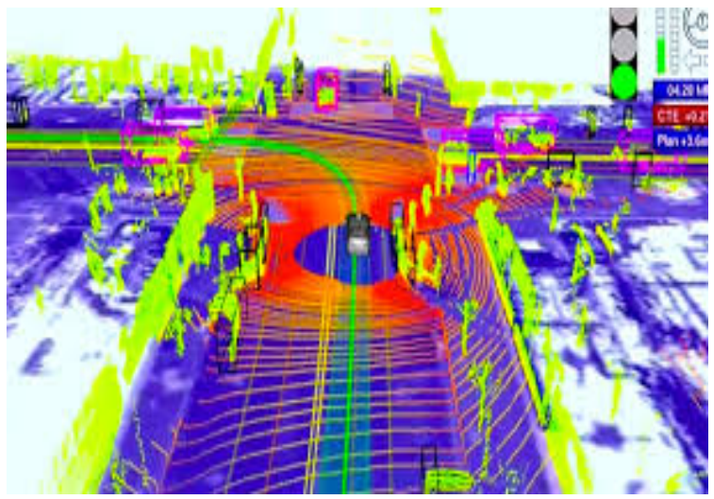
LIDAR
(905 nm, 1550 nm)



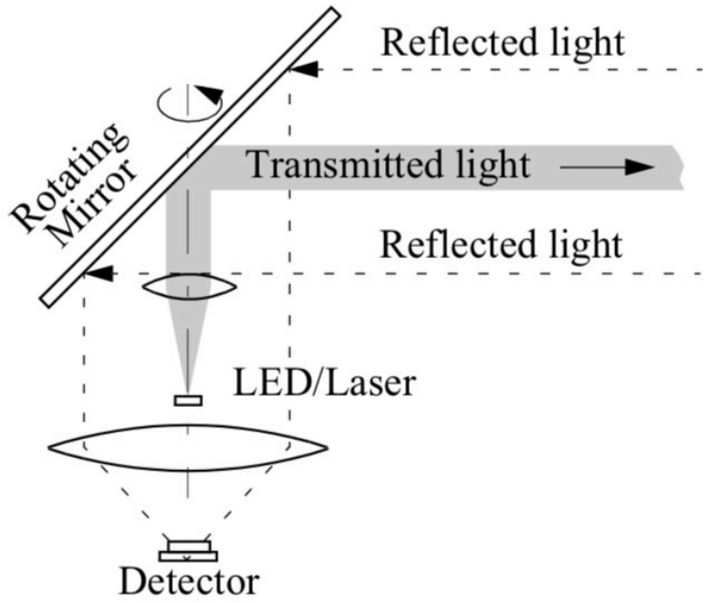
X-ray image

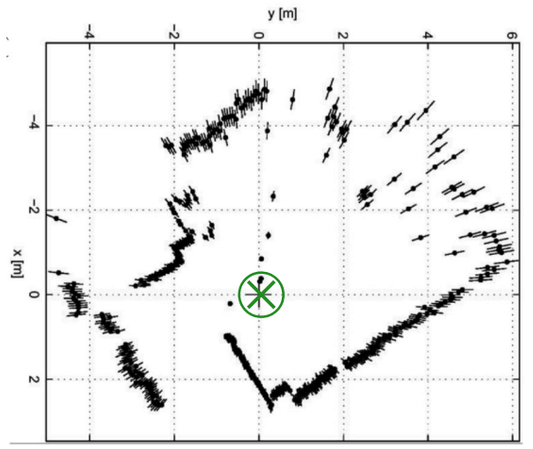
Light Detection and Ranging (LIDAR). LIDAR uses laser and calculates the time of flight for distance measurement. It typically achieves centimeter accuracy at a distance of over 10 m. LIDAR vastly improved the accuracy of distance measurement in natural environments and is the key technology that opened up the era of autonomous driving.





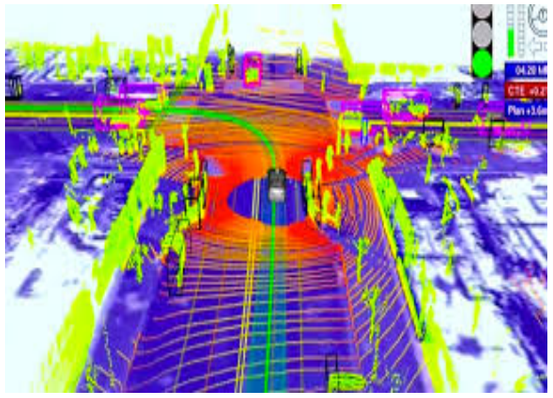
Sick LIDAR uses a single laser beam together with a rotating mirror to generate a 2-D scan, i.e., a curve in the 2-D space.





Velodyne LIDAR uses 64 laser beams on a rotating base to generate a 3-D scan, i.e., a surface in the 3-D space.

In principle, we can use a single beam to scan a 3-D surface sequentially, using a suitably constructed rotating platform. However, that takes substantial time per scan, resulting in a slow sensor.



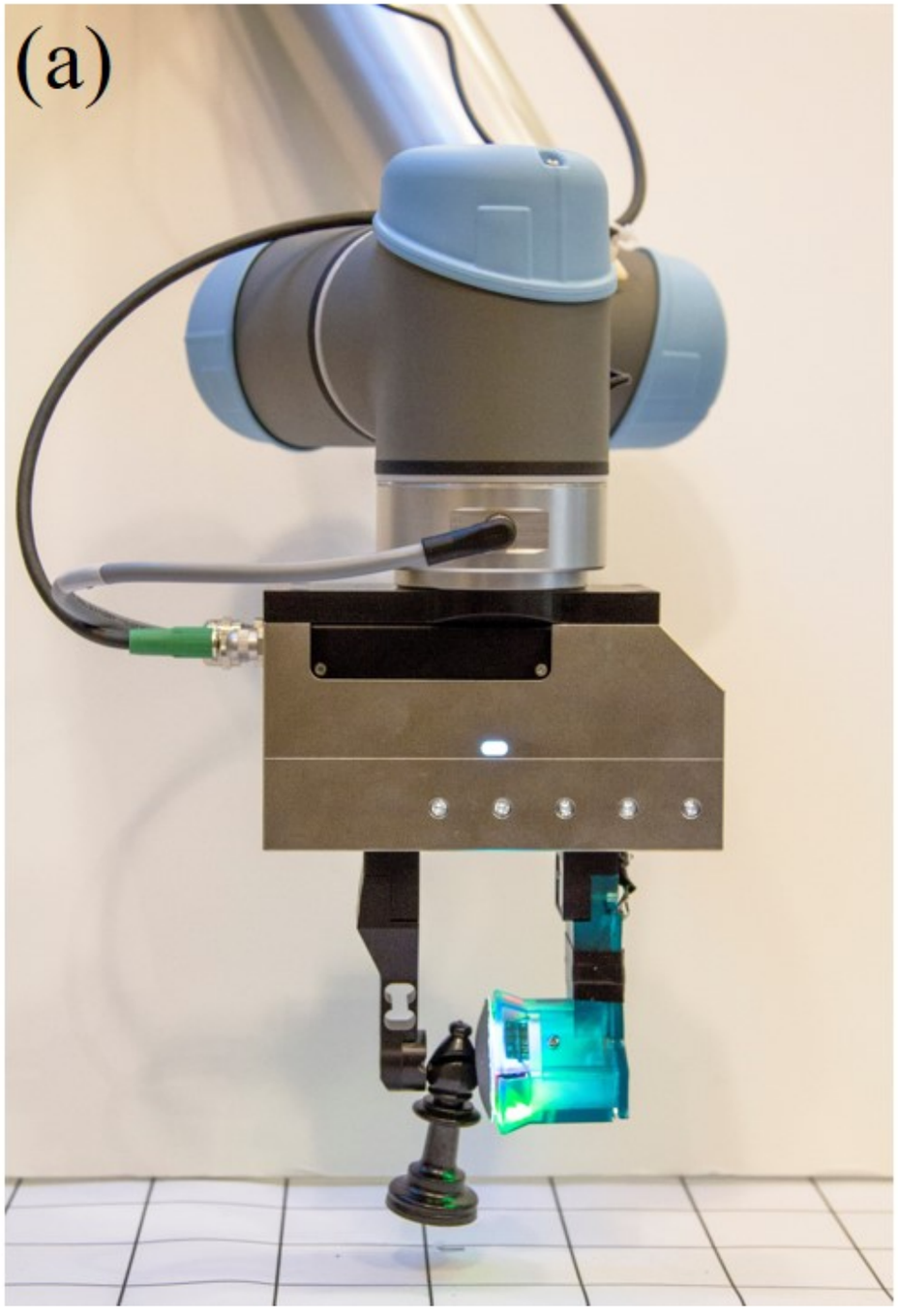
Ultrasonic distance sensors. Like LIDAR, the ultrasonic distance sensor calculates the time of flight for distance measurement, but it emits ultrasound, instead of laser.

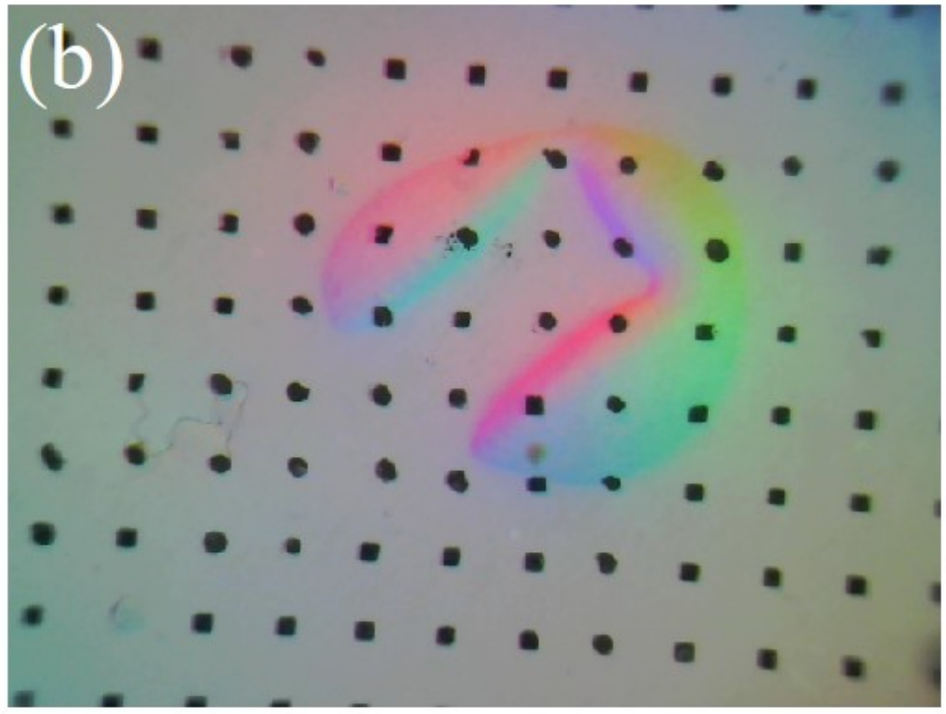
For comparison, LIDARs have much longer maximum range and higher accuracy. Ultrasonic distance sensors are smaller, lighter, and much cheaper.

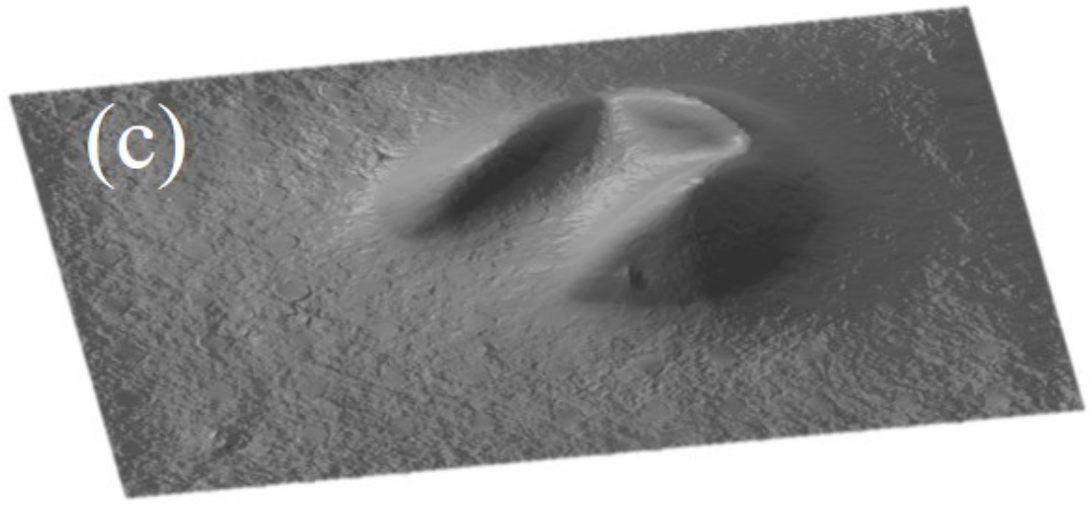
IR distance sensors. The IR distance sensor emits IR light and uses triangulation.

Touch Sensor: Gelsight

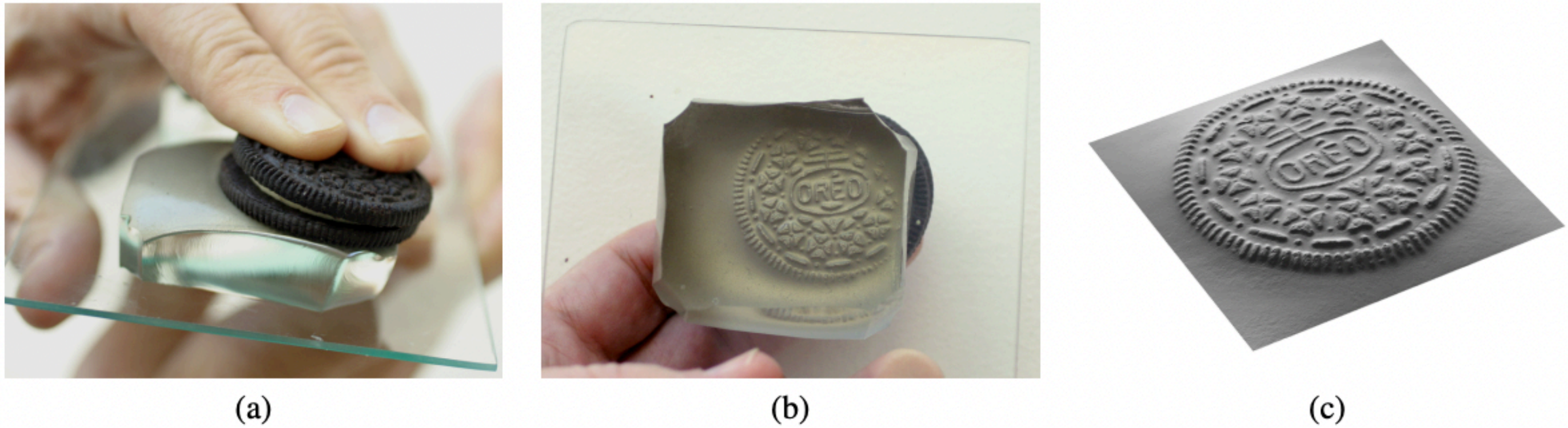
Turns a touch signal into an image







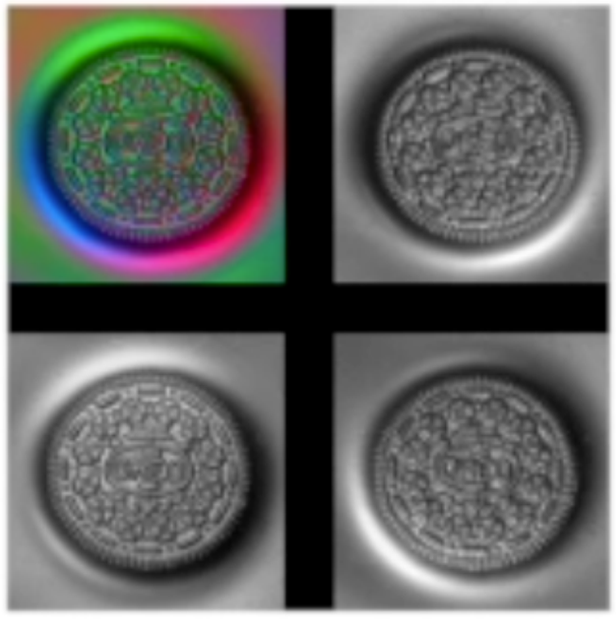
Touch Sensor: Gelsight



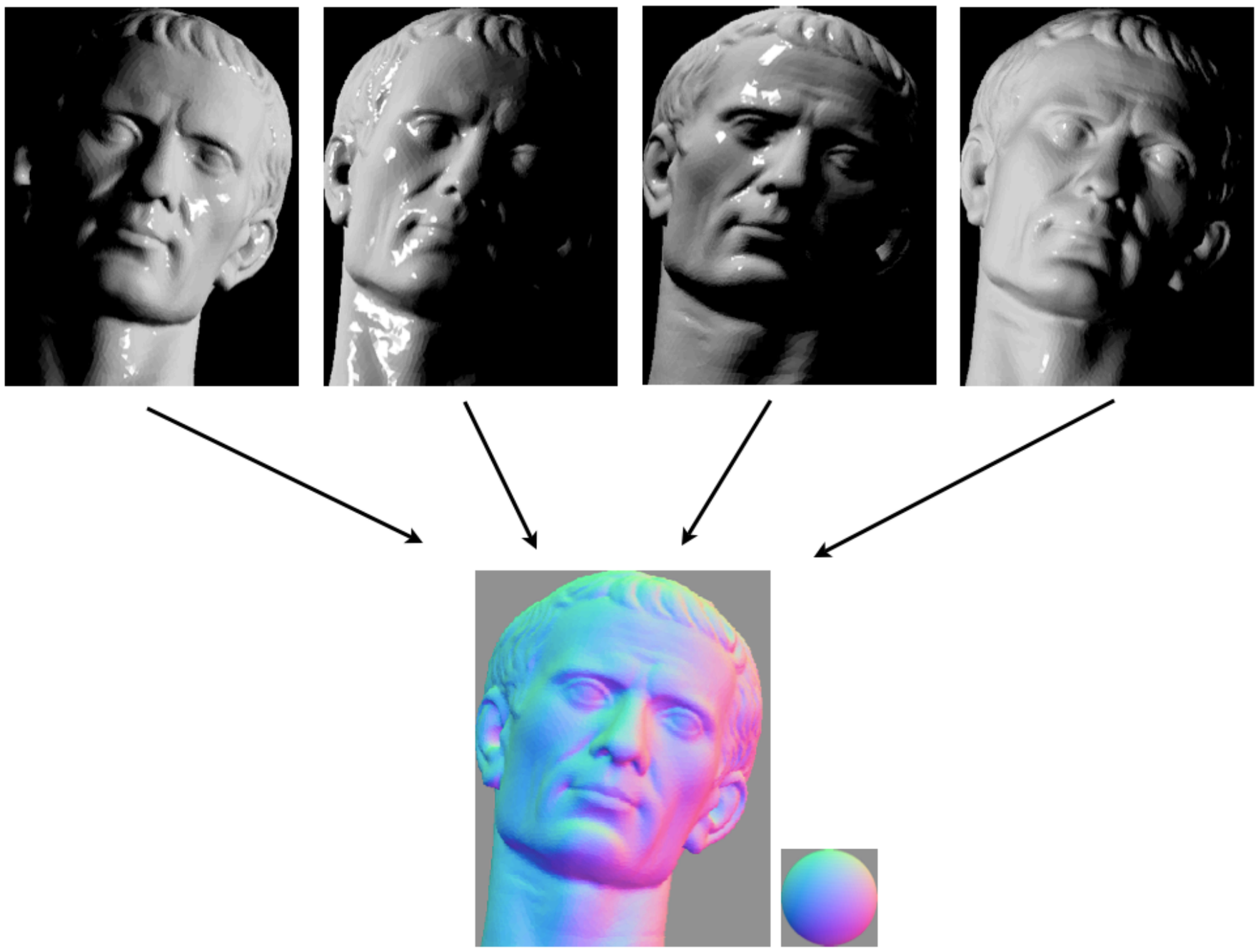
- Slab of clear elastomer
- Reflective “Skin” (membrane)
- Membrane deformation yields a shaded image
- Computer vision can estimate shape, etc.

Touch Sensor: Gelsight

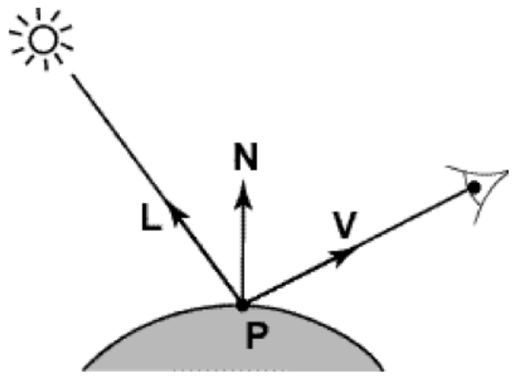
- Slab of clear elastomer
- Reflective "Skin" (membrane)
- Membrane deformation yields a shaded image
- Computer vision can estimate shape, etc.

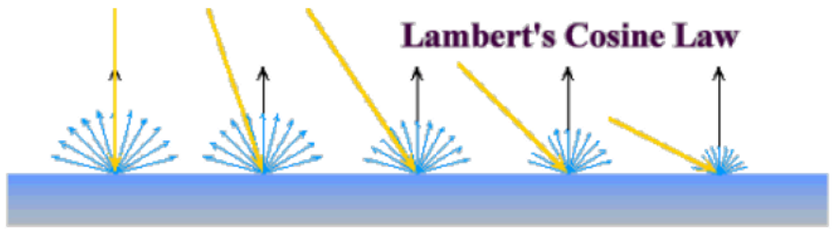


Photometric Stereo



Diffuse reflection





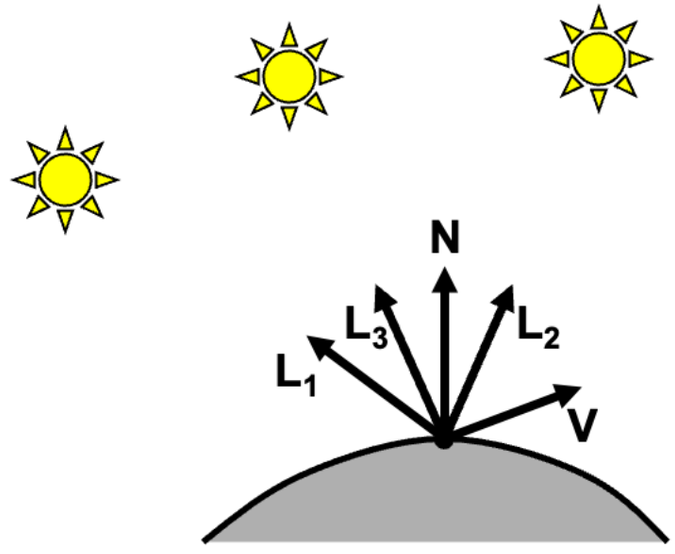
$$R_e = k_d \mathbf{N} \cdot \mathbf{L} R_i$$

image intensity of **P** → $I = k_d \mathbf{N} \cdot \mathbf{L}$

Simplifying assumptions

- I = R_e: camera response function f is the identity function:
  - can always achieve this in practice by solving for f and applying f^{-1} to each pixel in the image
- R_i = 1: light source intensity is 1
  - can achieve this by dividing each pixel in the image by R_i

Photometric stereo



$$
\begin{aligned}
I_1 &= k_d \mathbf{N} \cdot \mathbf{L_1} \\
I_2 &= k_d \mathbf{N} \cdot \mathbf{L_2} \\
I_3 &= k_d \mathbf{N} \cdot \mathbf{L_3}
\end{aligned}
$$

Can write this as a matrix equation:

$$
\begin{bmatrix} I_1 & I_2 & I_3 \end{bmatrix} = k_d \mathbf{N}^T \begin{bmatrix} \mathbf{L_1} & \mathbf{L_2} & \mathbf{L_3} \end{bmatrix}
$$

Solving the equations

$$\underbrace{\begin{bmatrix} I_1 & I_2 & I_3 \end{bmatrix}}_{\mathbf{I}} = \underbrace{k_d \mathbf{N}^T}_{\mathbf{G}} \underbrace{\begin{bmatrix} \mathbf{L_1} & \mathbf{L_2} & \mathbf{L_3} \end{bmatrix}}_{\mathcal{L}}$$

$1 \times 3 \qquad 1 \times 3 \qquad 3 \times 3$

$$\mathbf{G} = \mathbf{I}\mathbf{L}^{-1}$$

$$k_d = \|\mathbf{G}\|$$

$$\mathbf{N} = \frac{1}{k_d}\mathbf{G}$$

Poission Surface Reconstruction

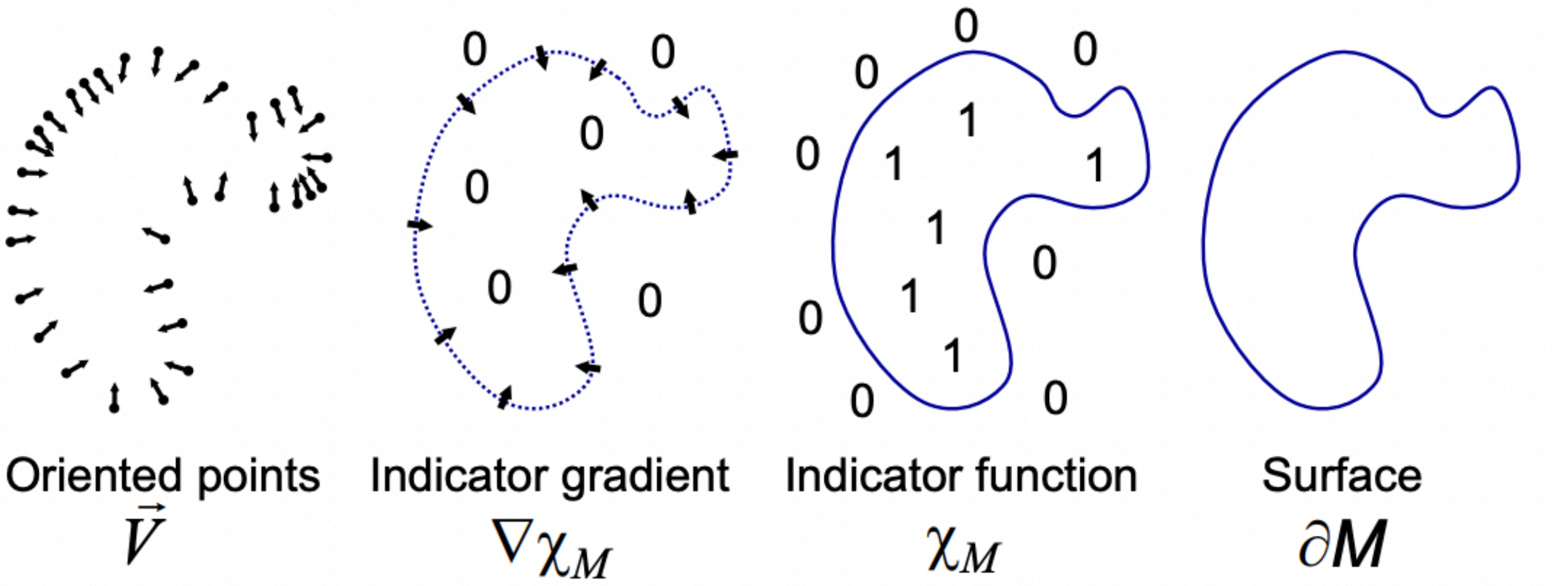
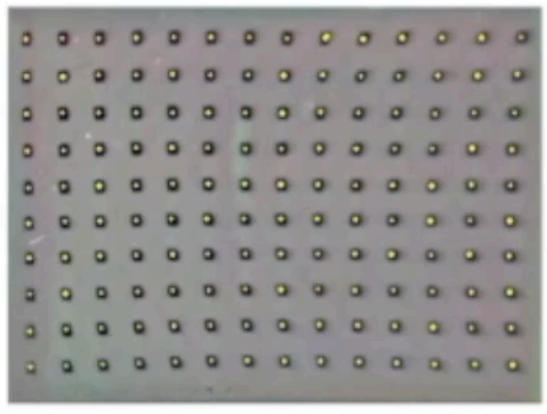


Figure 1: *Intuitive illustration of Poisson reconstruction in 2D.*

$$\min_{\chi} \left\| \nabla\chi - \vec{V} \right\|$$

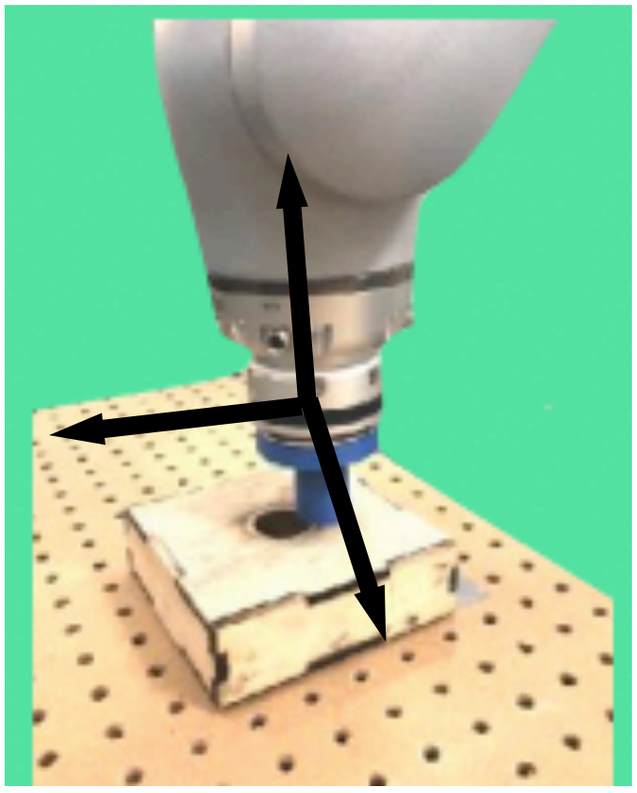
Touch Sensor: Gelsight

Markers

Sensor	Marker Displacement
	

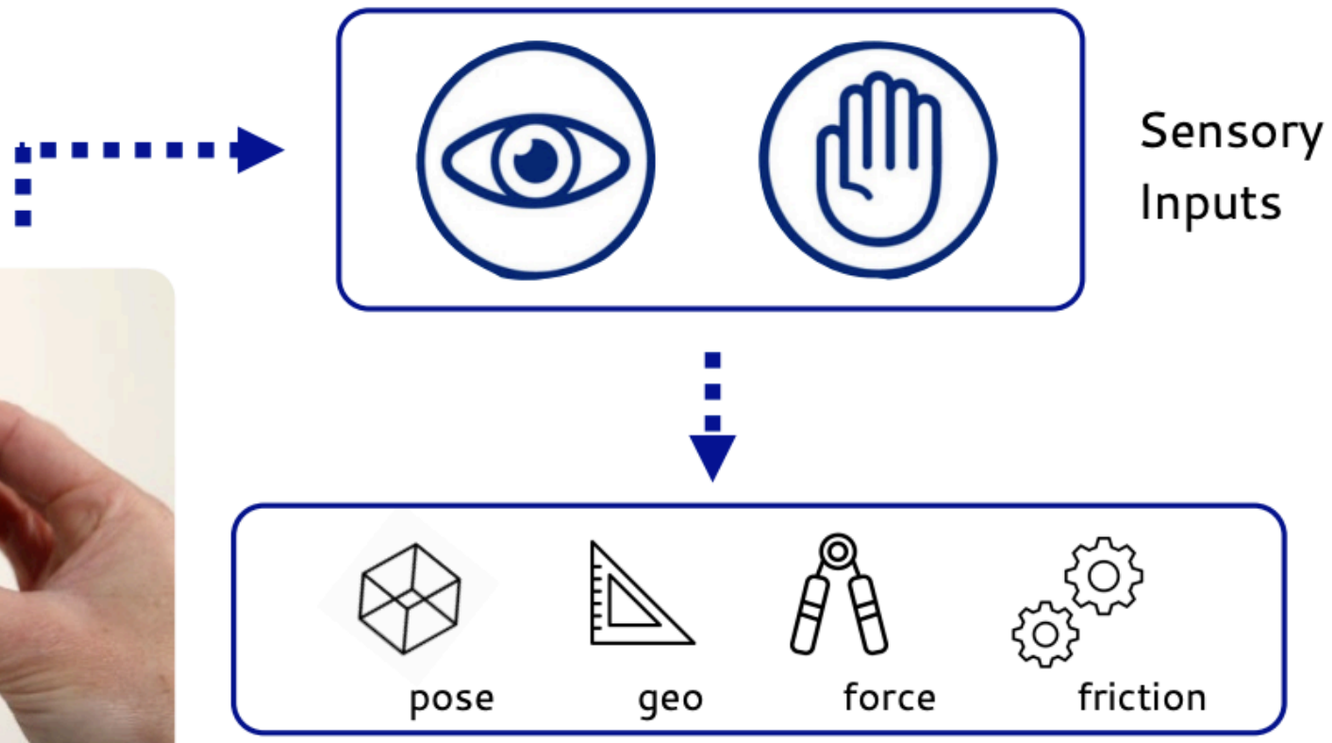
We can add the markers on the gel to track the marker displacement,

Force and Torque Sensor



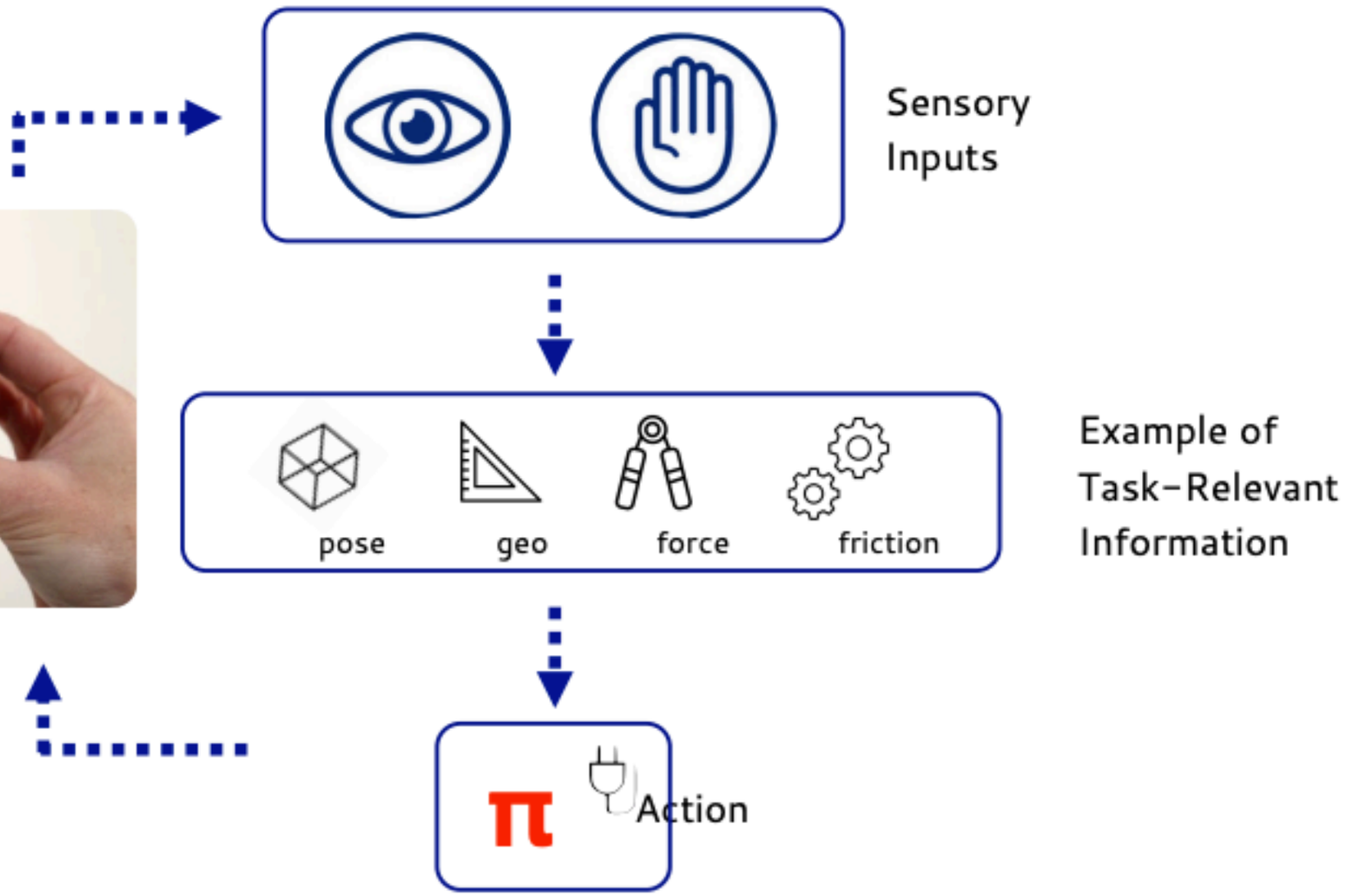
- A 6-axis Force Torque sensor measures forces and torques along the x-y-z axis

Human multimodal sensor-motor coordination

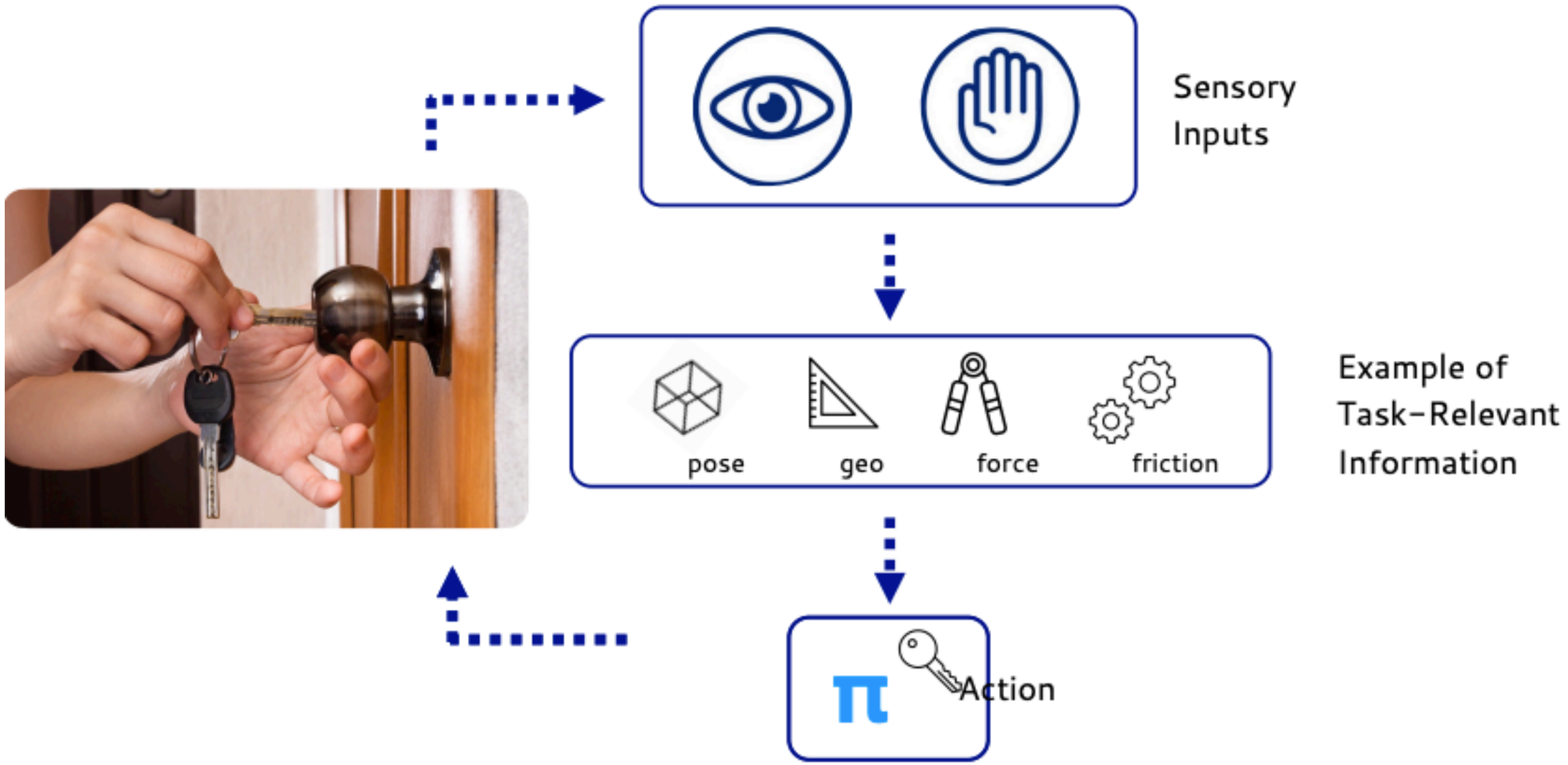


Example of
Task-Relevant
Information

Human multimodal sensor-motor coordination



Human multimodal sensor-motor coordination

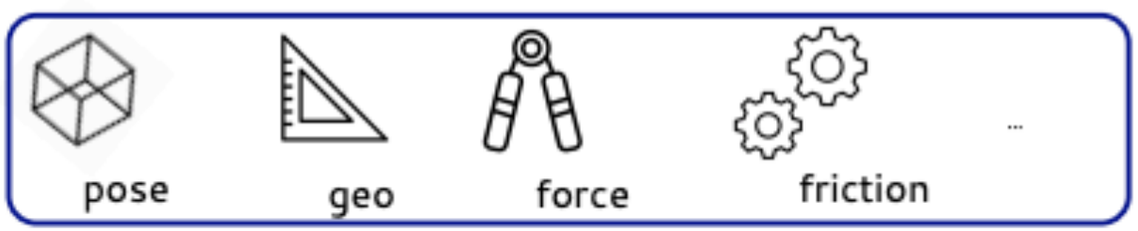


Generalizable representation for multimodal input

$$\pi(f(o_1, o_2, o_3 \ldots o_n)) = a$$

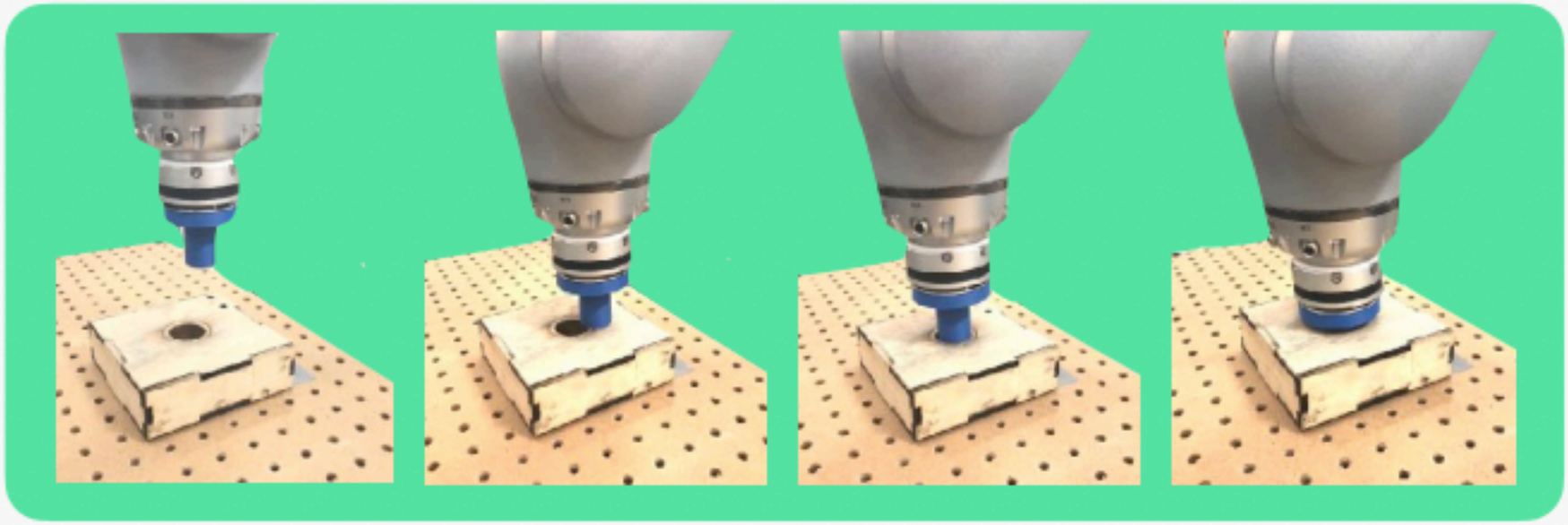
multimodal sensory inputs

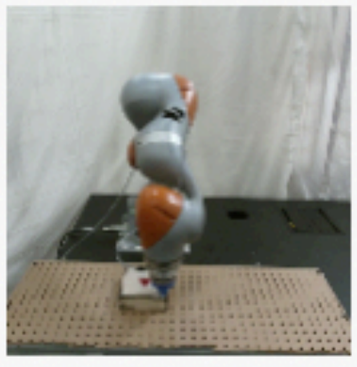
Learn representation



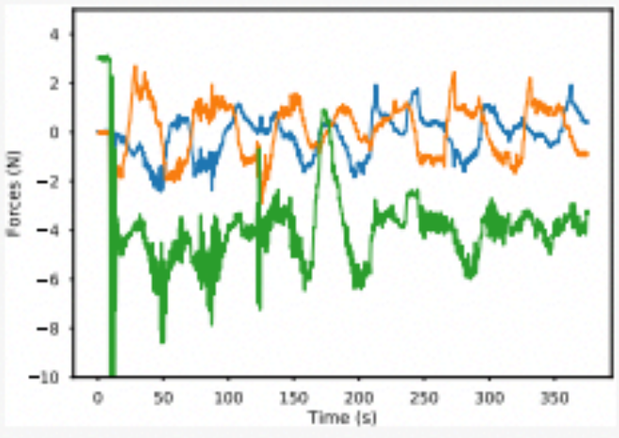
learn policy
for new task instances

Sensor Fusion



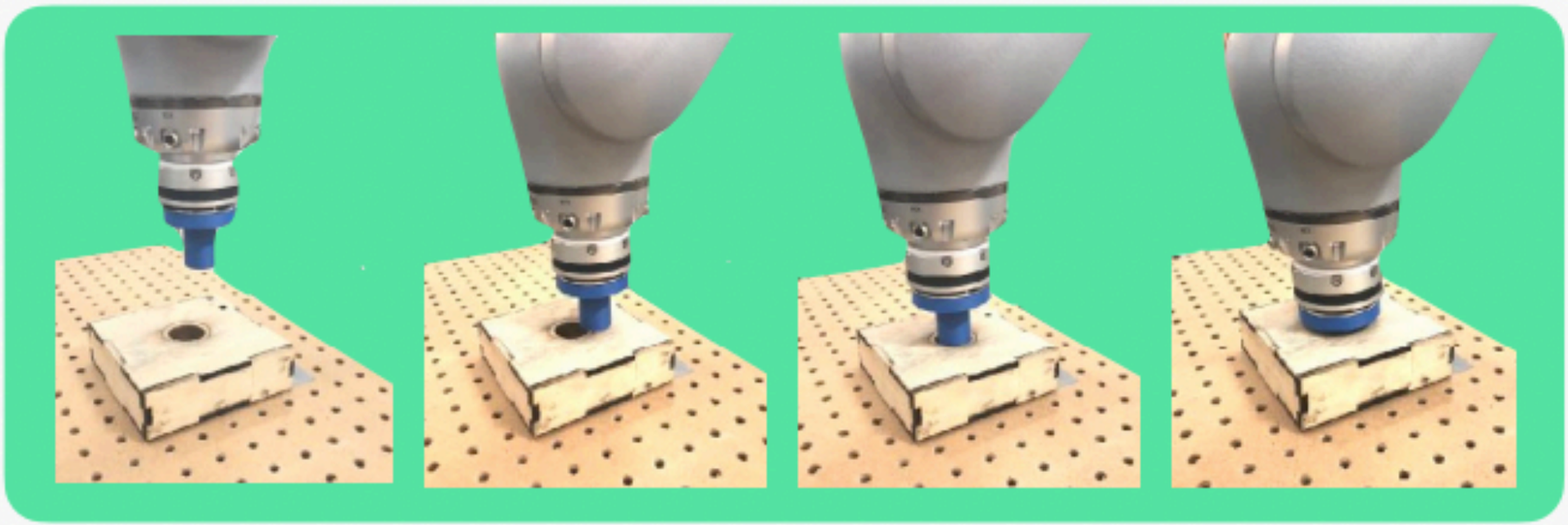


RGB



force/torque

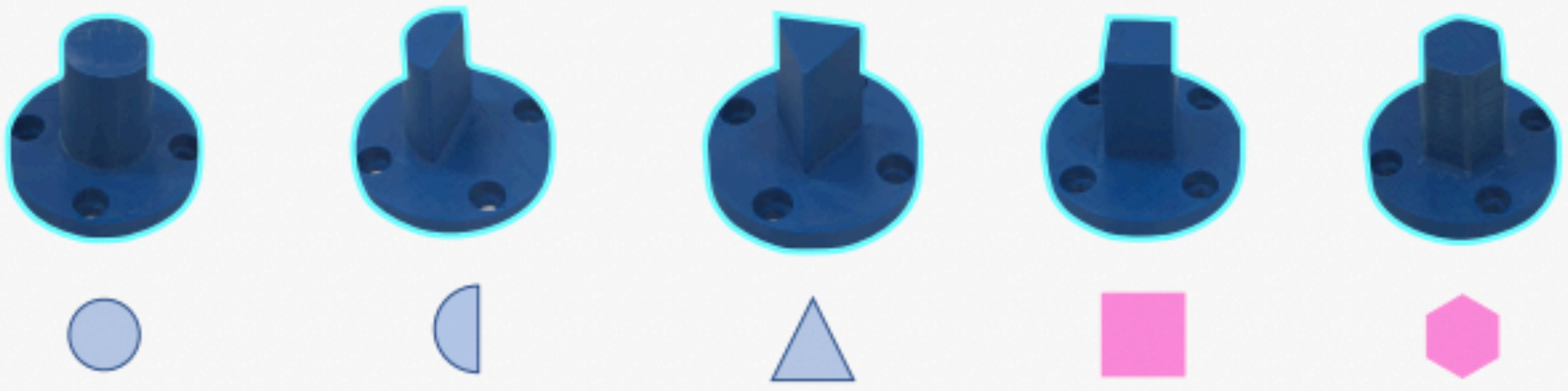
robot states

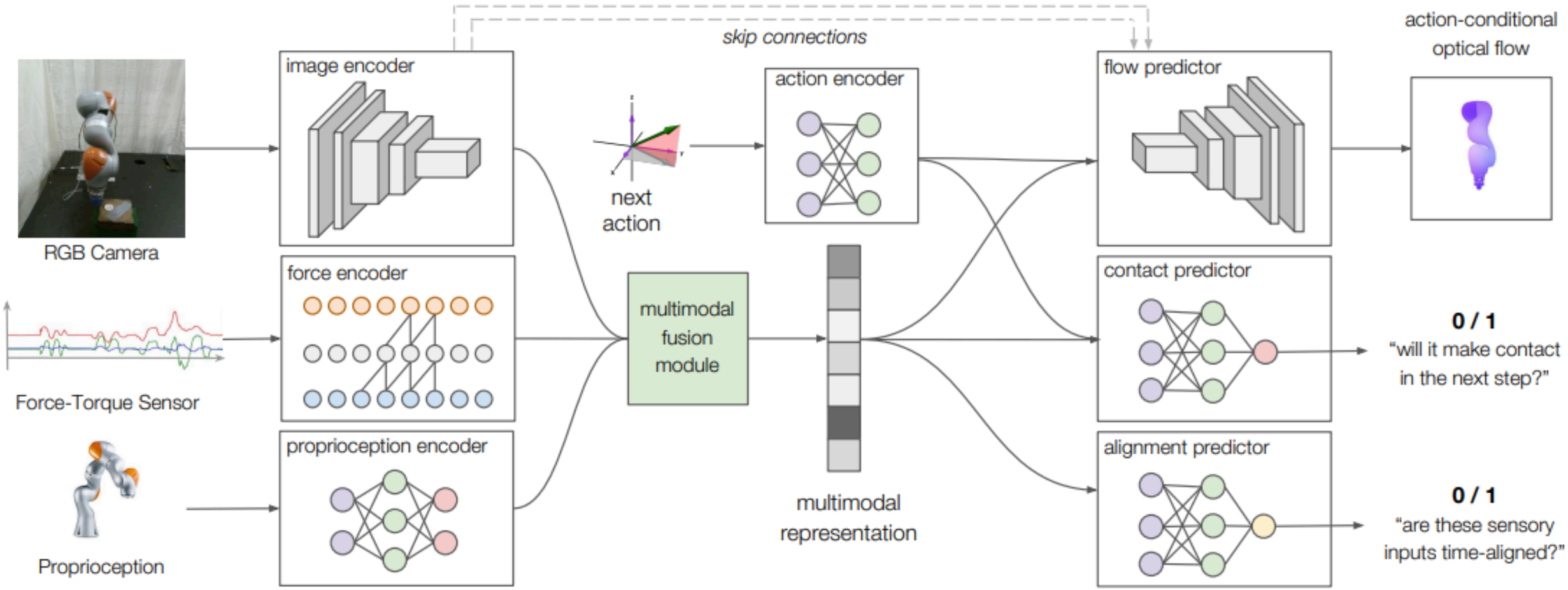


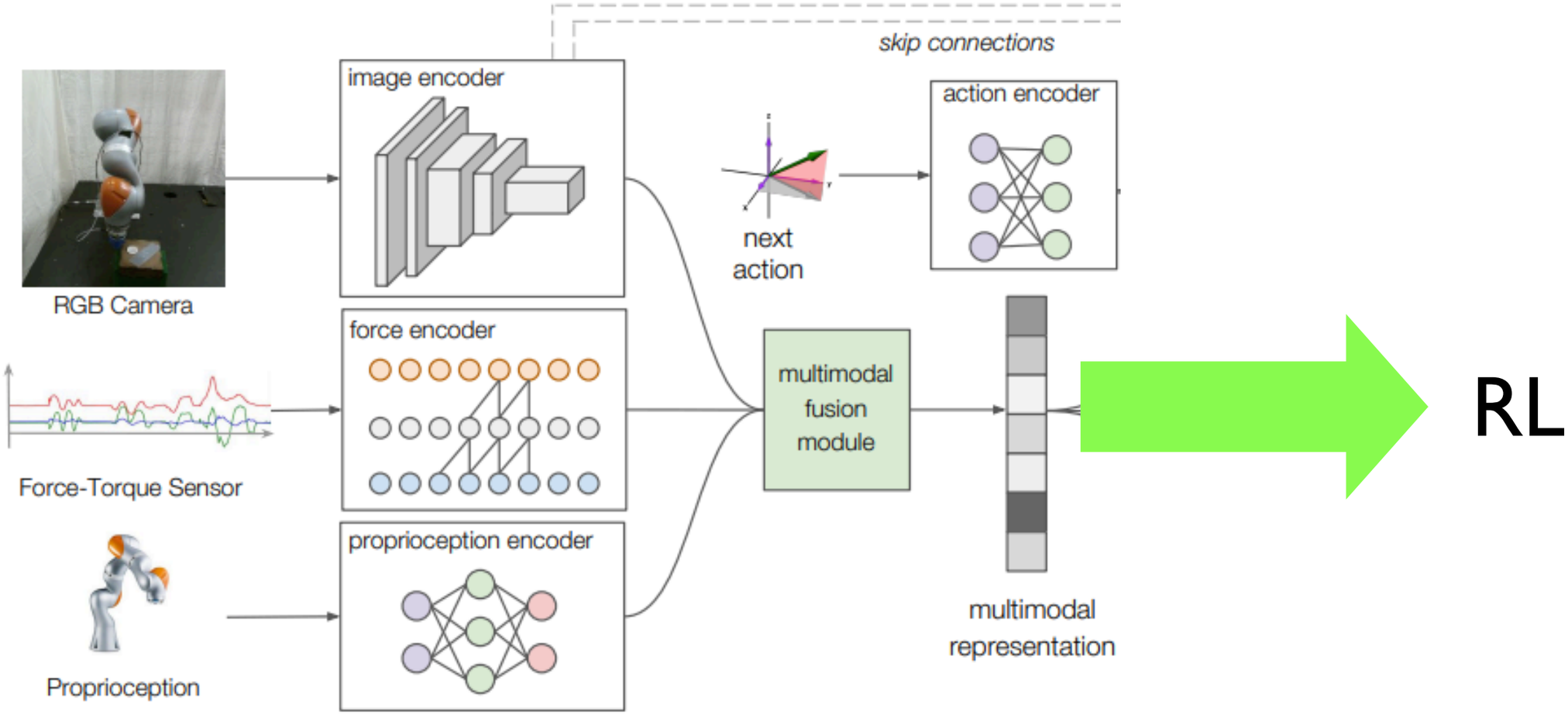
Training

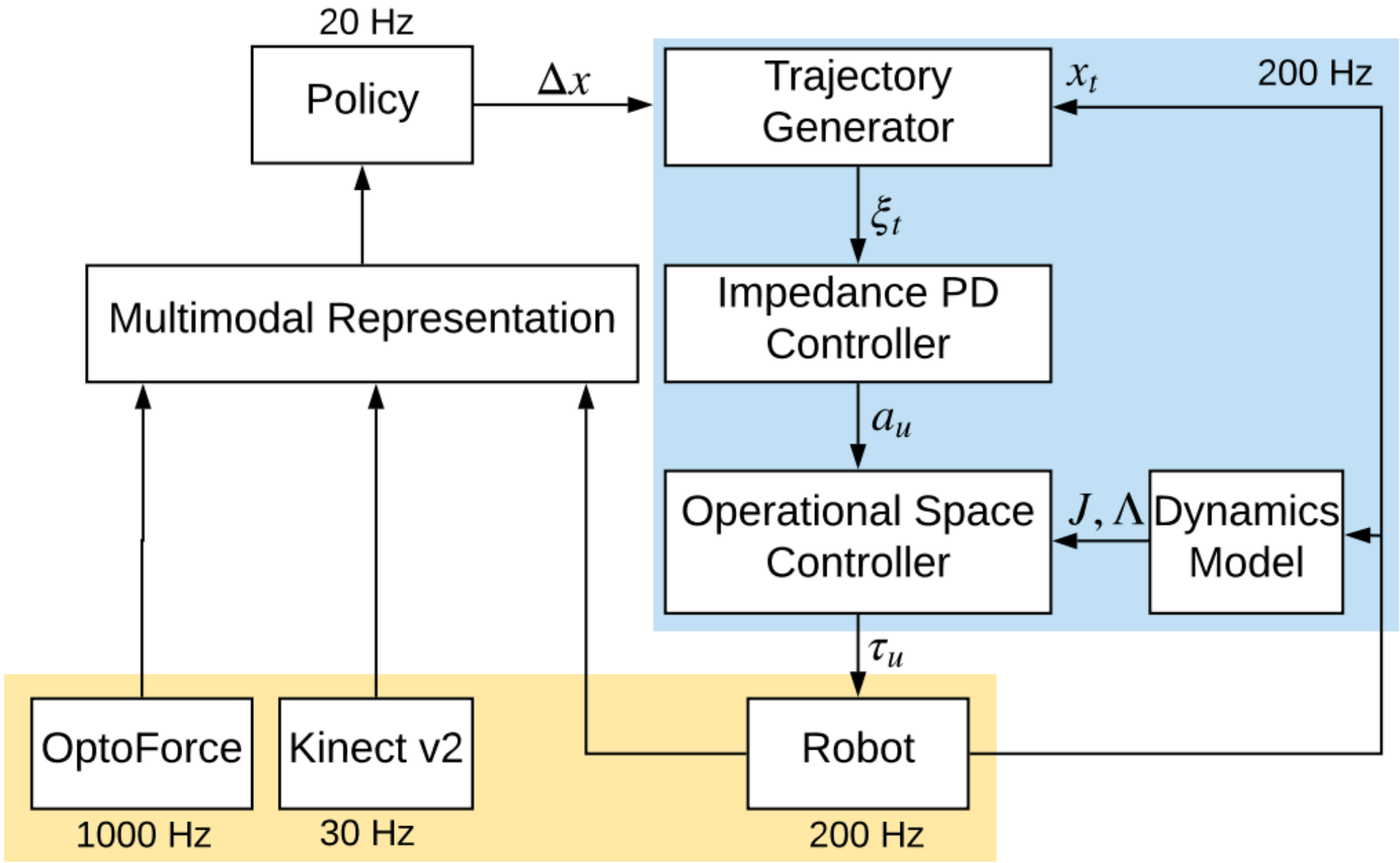
Testing

Peg geometry

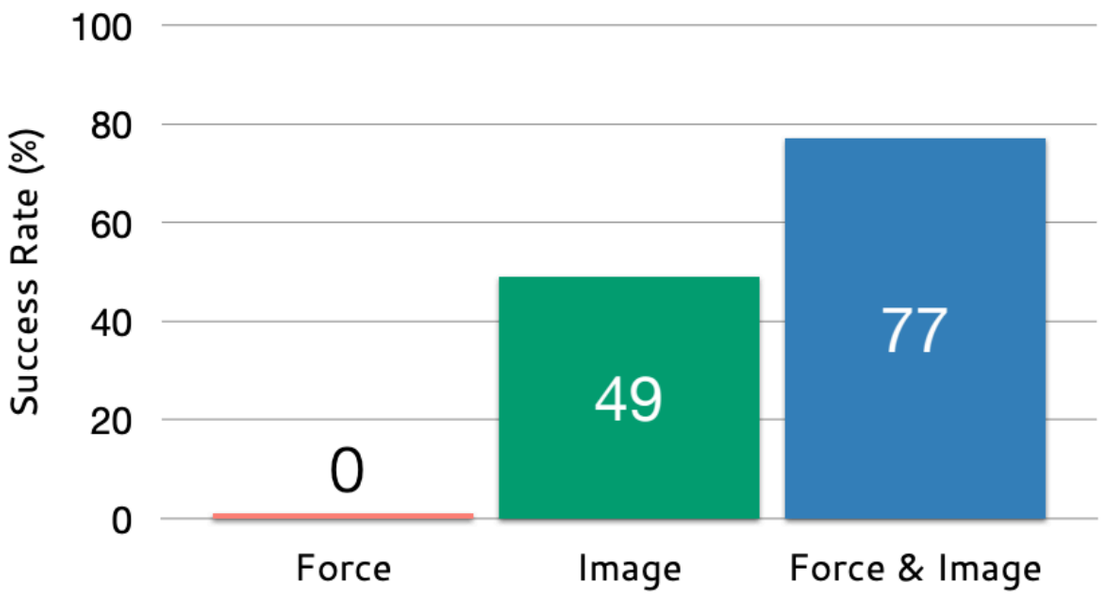



skip connections
image encoder
action encoder
flow predictor
action-conditional
optical flow
RGB Camera
next
action
force encoder
multimodal
fusion
module
contact predictor
0 / 1
"will it make contact
in the next step?"
Force-Torque Sensor
proprioception encoder
multimodal
representation
alignment predictor
0 / 1
"are these sensory
inputs time-aligned?"
Proprioception


skip connections
image encoder
action encoder
next
action
RGB Camera
force encoder
multimodal
fusion
module
Force-Torque Sensor
proprioception encoder
multimodal
representation
Proprioception
RL


20 Hz
Policy
Δx
Trajectory Generator
x_t
200 Hz
ξ_t
Impedance PD Controller
a_u
Operational Space Controller
J, Λ
Dynamics Model
τ_u
Multimodal Representation
OptoForce
1000 Hz
Kinect v2
30 Hz
Robot
200 Hz

How is each modality used?



Simulation Results
(Randomized box location)

Force Only: Can't find box

Image Only: Struggles with peg alignment

Force & Image: Can learn full task completion

Does the representation generalize?

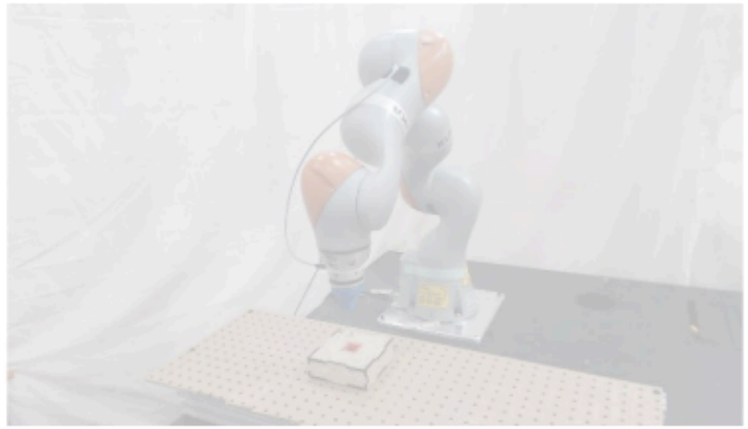
92% Success Rate

Tested on ▲

Representation ▲

Policy ▲

62% Success Rate



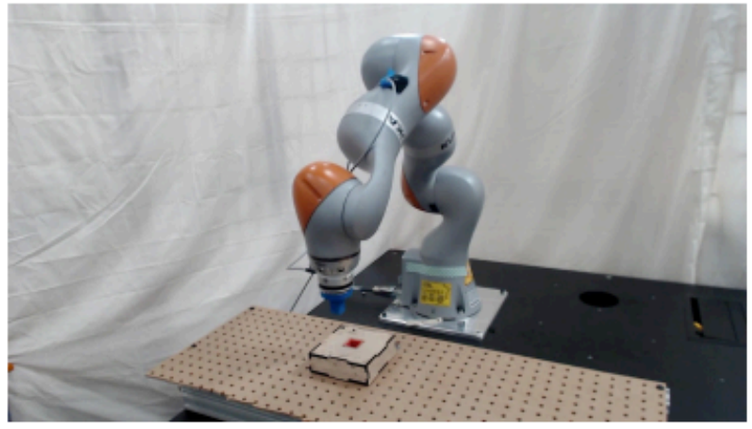
Policy does not transfer

Tested on ■

Representation ▲

Policy ▲

92% Success Rate



Representation transfers

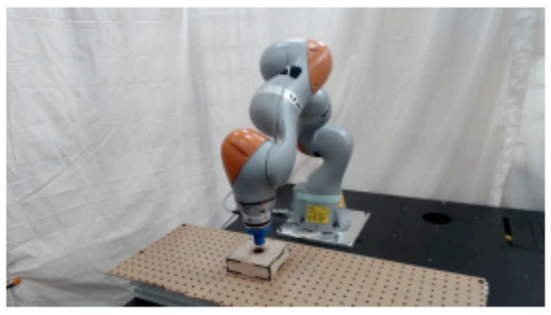
Tested on ■

Representation ▲

Policy ■

Self-supervised data collection

$o_{RGB}, o_{force}, o_{robot}$



100k data points
90 minutes

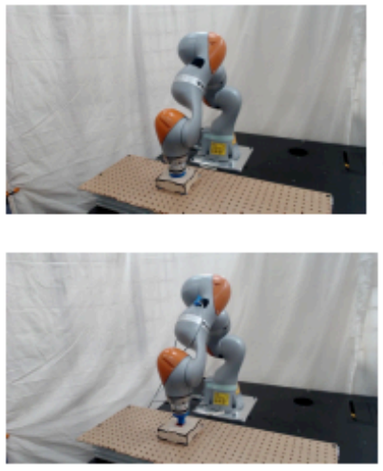
Representation learning

$f(o_{RGB}, o_{force}, o_{robot})$

20 epochs on GPU
24 hours

Policy learning

$\pi(f(\cdot)) = a$



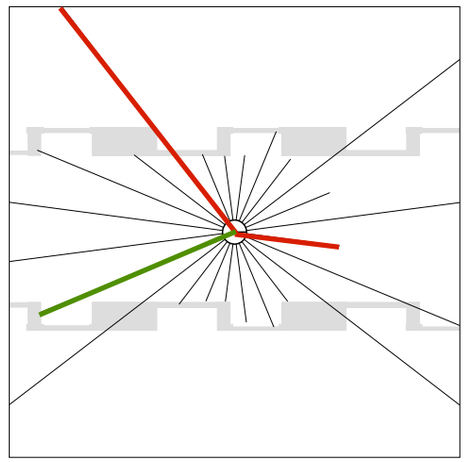
Deep RL
5 hours

Lee et al. Making Sense of Vision and Touch: Self-Supervised Learning of Multimodal Representations for Contact-Rich Tasks. ICRA'19. *Best Paper Award.*

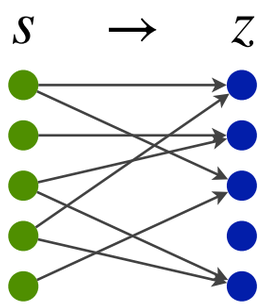
All sensors are noisy! Some are less noisy than others.

Sensor models. Let z be the measured distance. What is the actual distance s? The simplest answer is $s = z$. Unfortunately, that is often wrong, because sensors are noisy:

$$z = s + \delta\,.$$



Can we then find a function f that gives us $s = f(z)$? That turns out difficult, too. The mapping from s to z is many-to-many and could be highly ambiguous:



We need a probabilistic function or more precisely a conditional probabilistic distribution $p(s \mid z)$ that gives the probability of s given z. How do we obtain such a probability distribution?

Model learning. Conceptually, $p(s \mid z)$ is a function of two variables. We discretize the s-z space into a set of bins. Given sufficient labelled data pairs (z_i, s_i), $i = 1,2,...,N$, the discriminative model directly estimates $p(s \mid z)$ by counting. This works if we have sufficient labelled data.

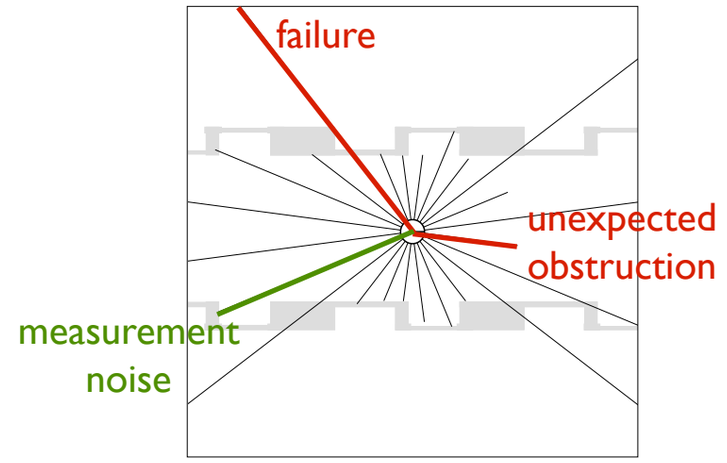
Alternatively, let us think about how z is "generated" from s as a result of noise. If we know $p(z \mid s)$, then we can apply the Bayes' rule to obtain $p(s \mid z)$:

$$p(s \mid z) = \eta p(z \mid s) p(s)$$

where η is a normalizing constant.

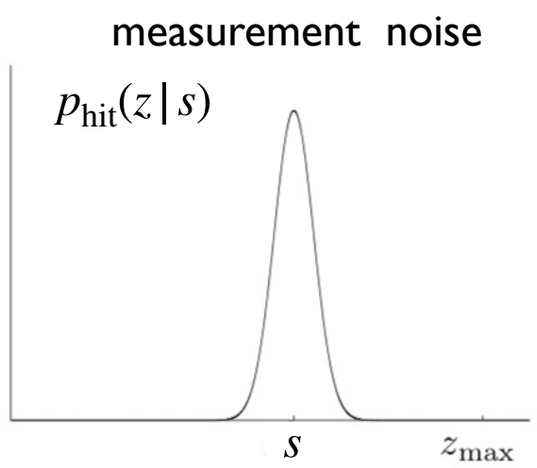
What are the different sources of error δ ?

- Measurement noise. The measured value is the true value s offset by a small amount that follows the Gaussian distribution with mean 0 and standard deviation σ_{hit}.
- Unexpected obstruction. The measured value is smaller than expected, because the beam is obstructed by an unexpected object. For example, the robot expects to measure the distance to the wall, but a person suddenly appears in between. One can work out that the shortened distance follows an exponential distribution with delay rate λ_{short}.

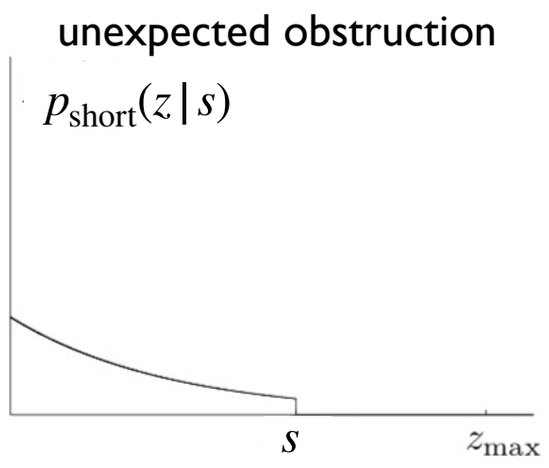


- Failures. The emitted beam does not reflect back at all, because of specular reflection by polished surfaces, total absorption by black, light-absorbing surface materials, … The sensor reports the maximum measurement value $z_{\max}$, while the obstruction could be much closer.
- Random measurements. This is "garbage collection". Occasionally the sensor reports erroneous values for any number of unexpected reasons with the sensor or the environment. We assume that the random measurement is distributed uniformly between 0 and $z_{\max}$.

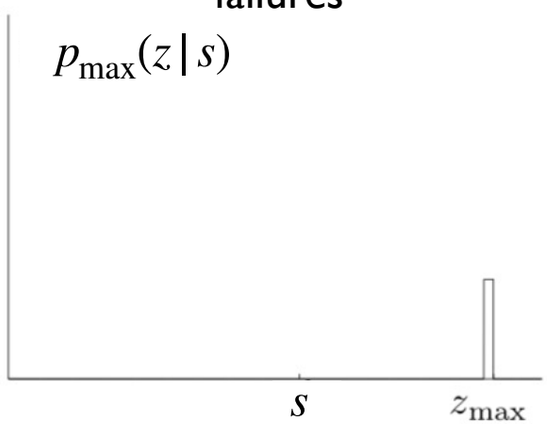
measurement noise



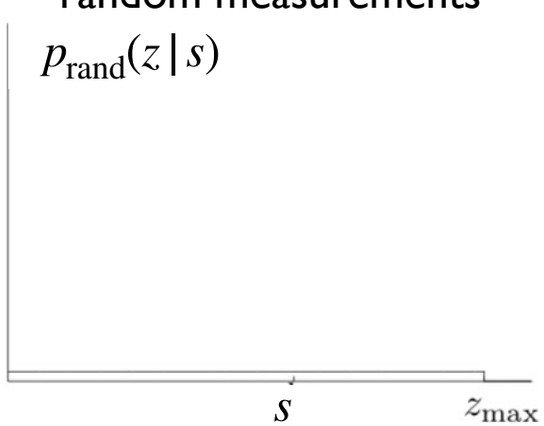
unexpected obstruction



failures



random measurements



Finally, $p(z \mid s)$ is a weighted sum of these four distributions:

$$p(z \mid s) = w_{\text{hit}}\, p_{\text{hit}}(z \mid s) + w_{\text{short}}\, p_{\text{short}}(z \mid s) + w_{\text{max}}\, p_{\text{max}}(z \mid s) + w_{\text{rand}}\, p_{\text{rand}}(z \mid s),$$

where the weights must sum up to 1.

Now given a dataset (z_i, s_i), $i = 1,2,...,N$, we just need to estimate a small number of parameters: the weights, plus σ_{hit} for p_{hit}, and λ_{short} for p_{short}.

Compare with the density estimation problem for $p(s \mid z)$, which requires estimating $M \times N$ parameters, if s and z take M and N discrete values, respectively. The generative model works if our assumptions on the measurement noise, obstruction, sensor failure, … are reasonably satisfied in practice.

Summary.

- Distance measurement. Active ranging sensors measure distance between the robot and objects in the environment. They differ in the emitted waves (sound, radio, light, …) and the principle of measurement (time-of-flight, triangulation, …).
- Passive sensors measure ambient environmental energy entering the sensor, e.g., accelerometers, cameras. Active sensors. Active sensors emit energy into the environment and then measure the environmental response, e.g., encoders or LIDAR.
- There are a wide variety of sensors. They differ in range, resolution, accuracy, bandwidth, physical dimensions, weight, and cost. Technical specifications report on these performance metrics, but **environmental factors** may substantially impact the actual performance.

Key concepts.

- Main sensor performance characteristics.
- IMU
- LIDAR
- Gelsight
- FT Sensor
- Sensor Fushion
- Sensor model $p(z \mid s)$