Robot Sensing

NA 568, Winter 2024

Minghan Zhu

Monday, February 19, 2024

Exteroceptive and proprioceptive sensors

- **Proprioceptive** sensors measure values internal to the system (robot); e.g. motor speed, wheel load, robot arm joint angles, battery voltage.
  - IMUs (accelerometers, gyroscopes), joint encoders, etc.
- **Exteroceptive** sensors acquire information from the robot’s environment; e.g. distance measurements, light intensity, sound amplitude.
  - Cameras, radars, lidars, etc.
- We focus on exteroceptive sensors in this lecture.

https://www.cs.cmu.edu/~rasc/Download/AMRobots4.pdf

Passive and active sensors

- **Passive** sensors measure ambient environmental energy entering the sensor.
  - CCD or CMOS cameras, microphones, etc.
- **Active** sensors emit energy into the environment, then measure the environmental reaction.
  - Lidars, radars, etc.
- Active sensors often achieve better performance but may cause and be subject to interference.

https://www.cs.cmu.edu/~rasc/Download/AMRobots4.pdf

Monocular cameras

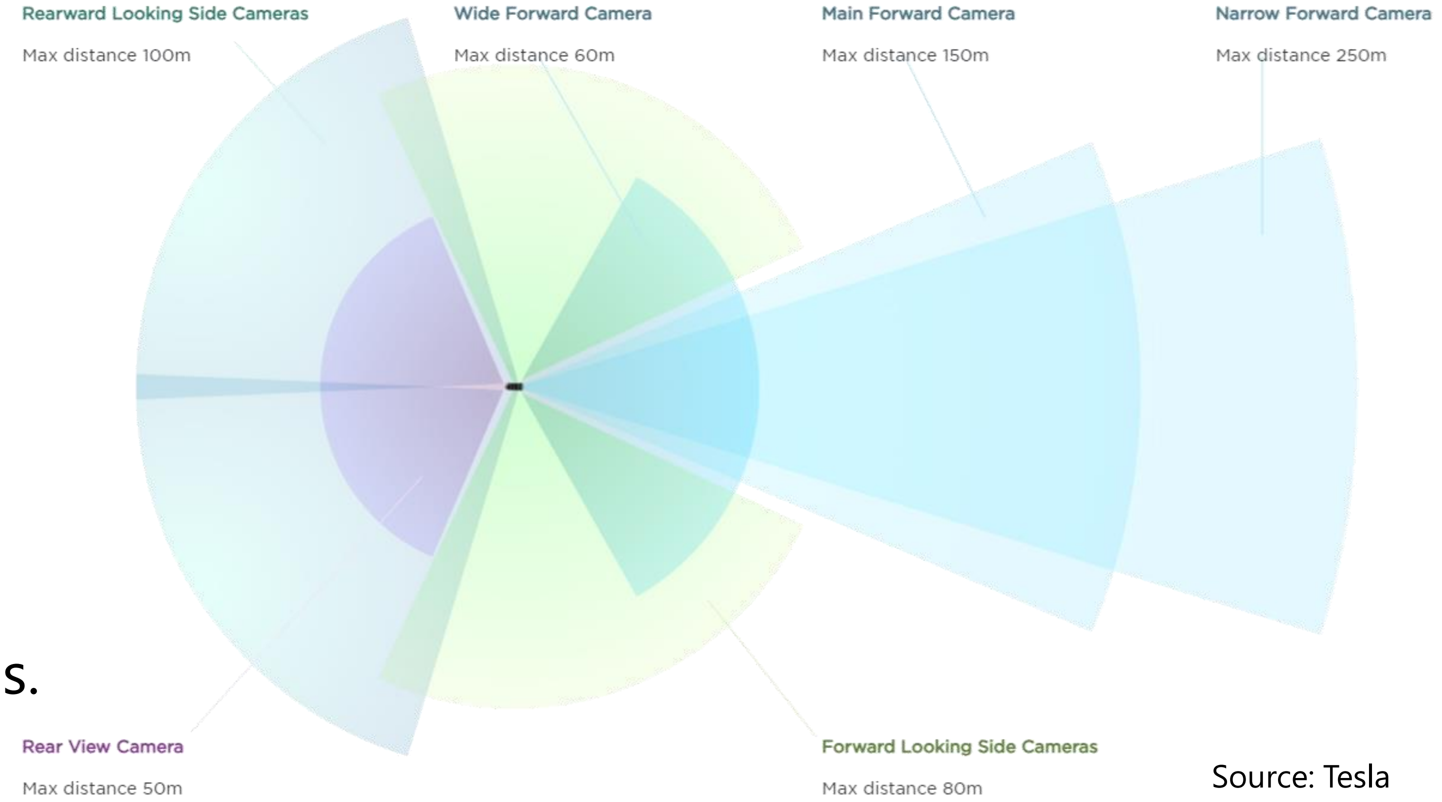
Camera (monocular)

- Receiving visible light ($\lambda \approx 380\sim740\ \mathrm{nm}$) from the environment.
- Strengths:
  - High resolution ($< 0.1°$).
  - Color information.
  - Affordable (<$100).
- Weaknesses:
  - No depth measurement.
  - Sensitive to lighting conditions.

Source: Tesla

Camera (monocular)

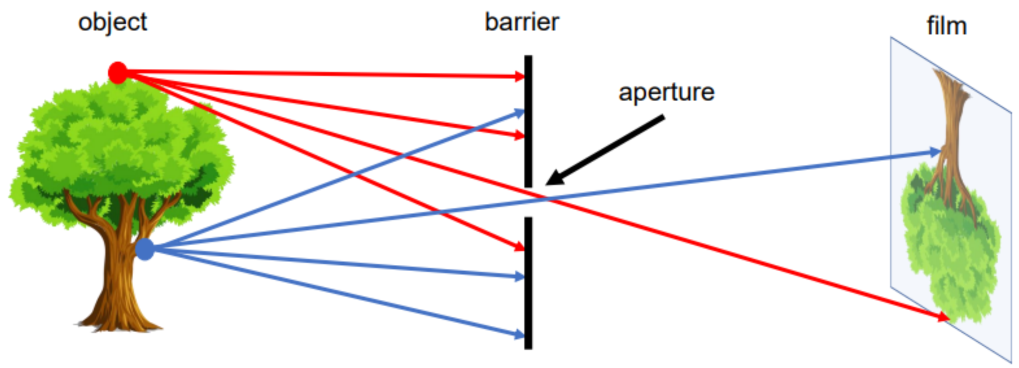
- Receiving visible light ($\lambda \approx 380{\sim}740\ \mathrm{nm}$) from the environment.
- Strengths:
  - High resolution ($< 0.1°$).
  - Color information.
  - Affordable (<$100).
- Weaknesses:
  - No depth measurement.
  - Sensitive to lighting conditions.

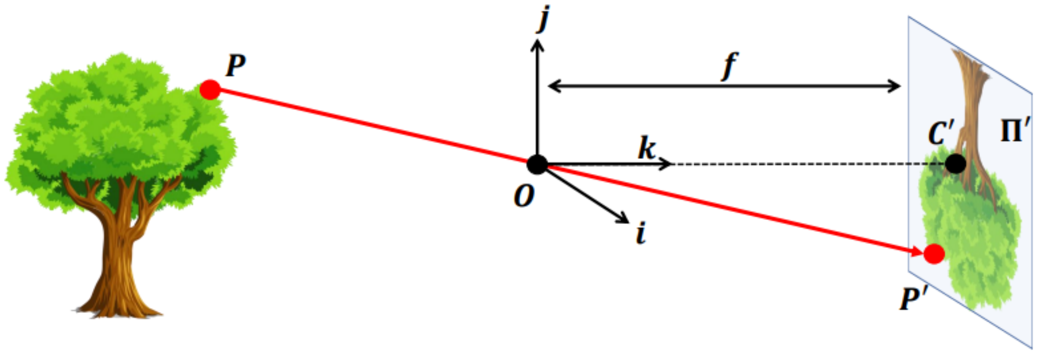


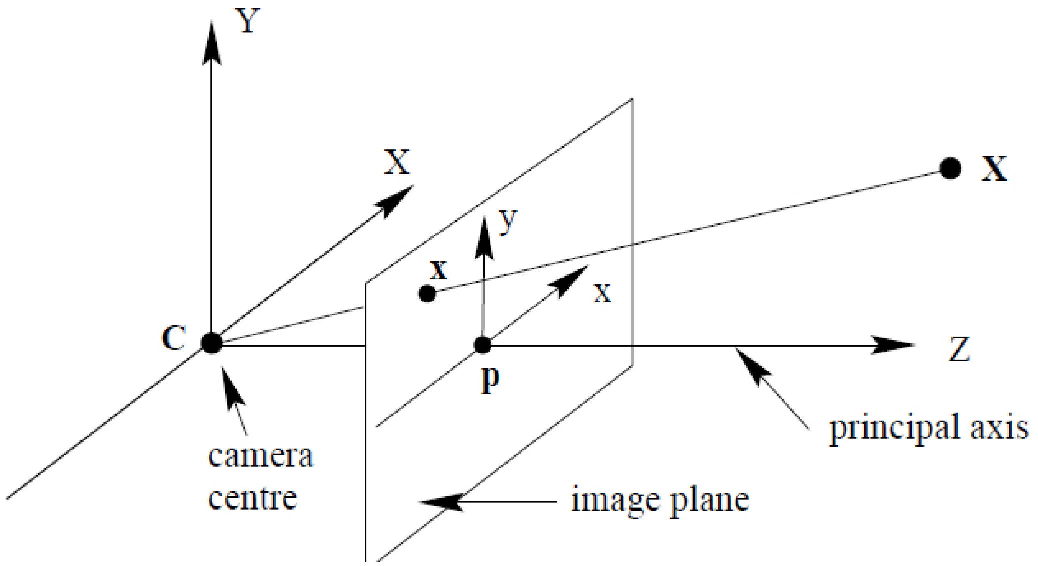
Source: Tesla

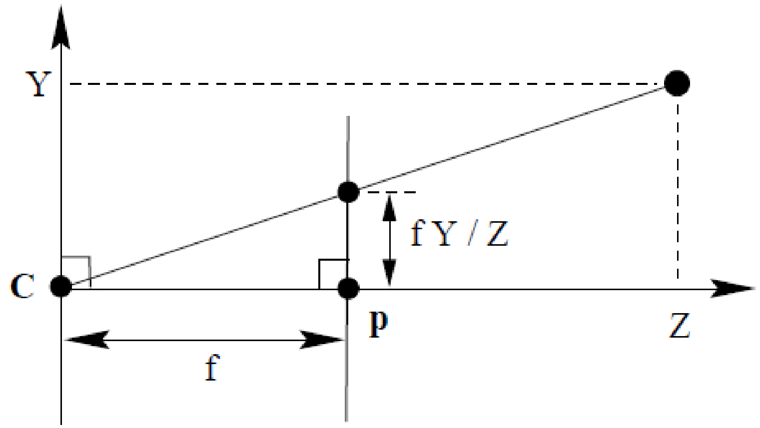
Pinhole camera model

- Simplified camera model







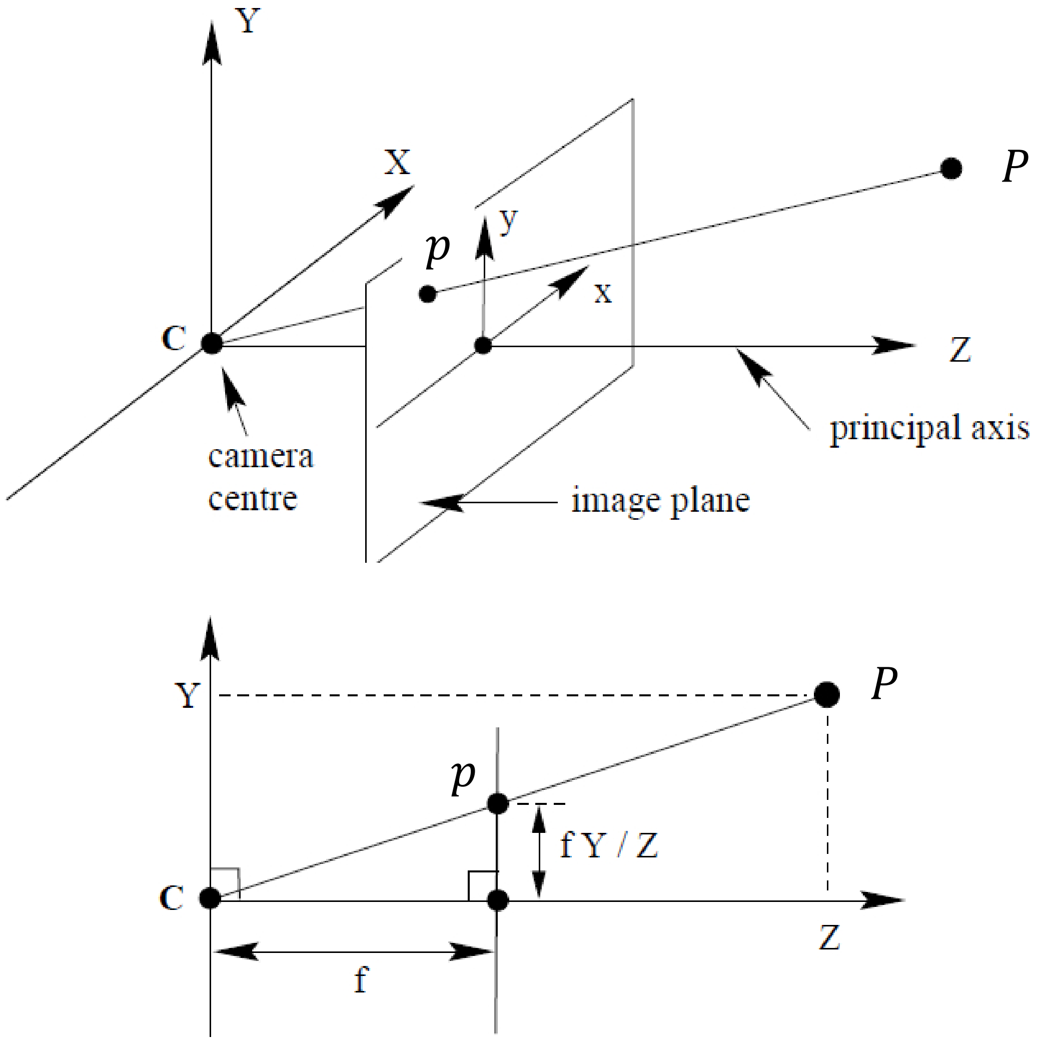


Source: Hata and Savarese

Source: HediVision

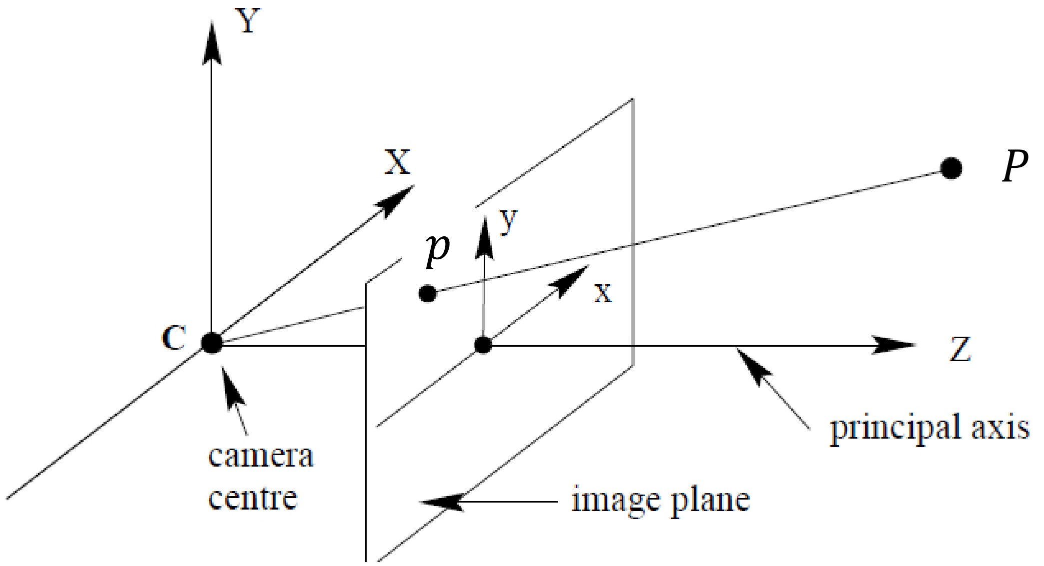
Pinhole camera model

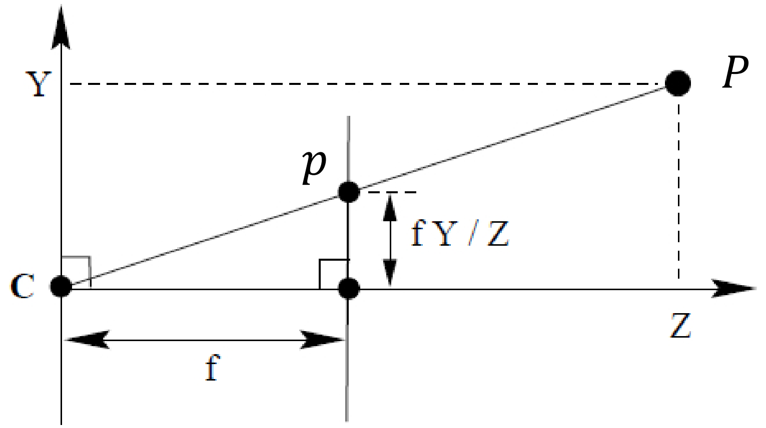
- Simplified camera model
- $P = (X, Y, Z)^T$
- $p = (x, y)^T$
- $y = fY/Z$
- $x = fX/Z$
- Having division is inconvenient. Can we write it in matrix form?



Pinhole camera model

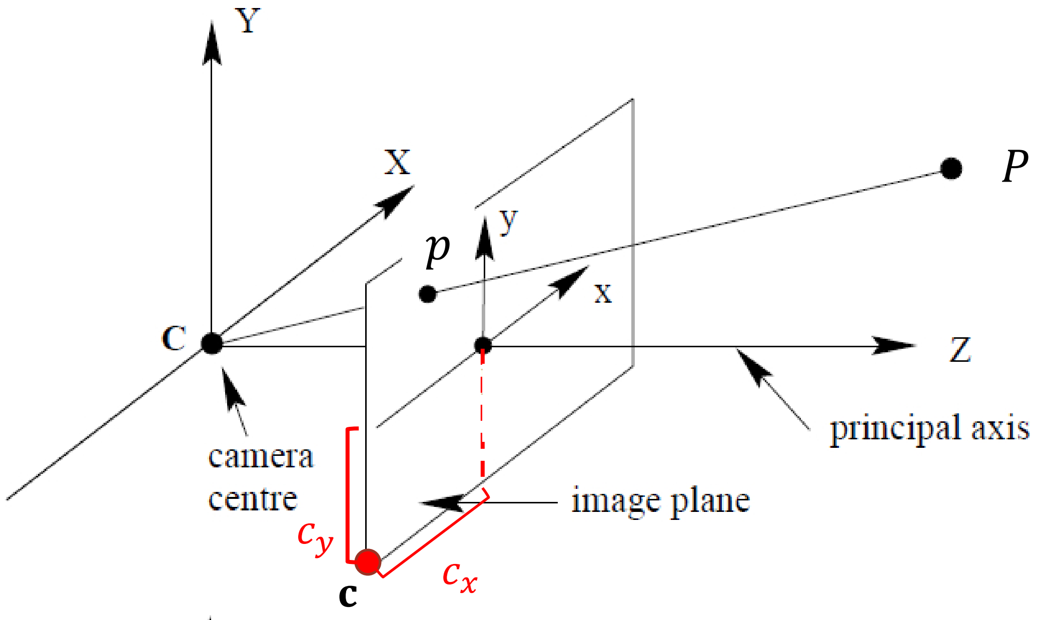
- $(X, Y, Z)^T \longmapsto (x, y)^T = (fX/Z, fY/Z)^T$
- Use homogeneous coordinate.
- $(x, y)^T \rightarrow (x, y, 1)^T$
- $(x, y, 1)^T = (\rho x, \rho y, \rho)^T, \forall \rho \neq 0$
- $\begin{pmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} fX \\ fY \\ Z \end{pmatrix} = \begin{pmatrix} fX/Z \\ fY/Z \\ 1 \end{pmatrix} = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$

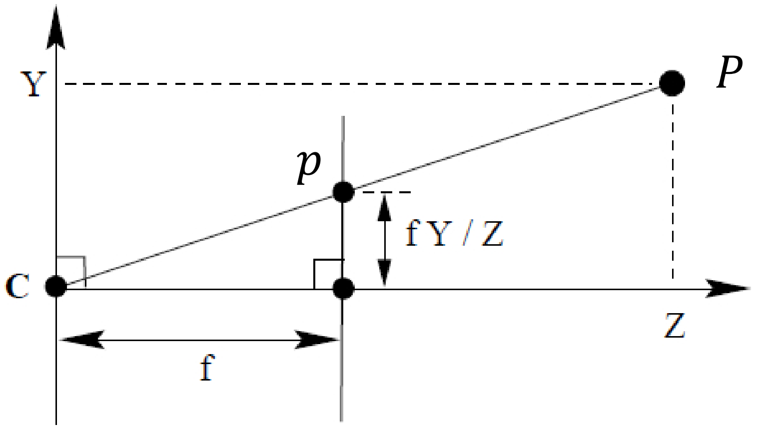




Pinhole camera model

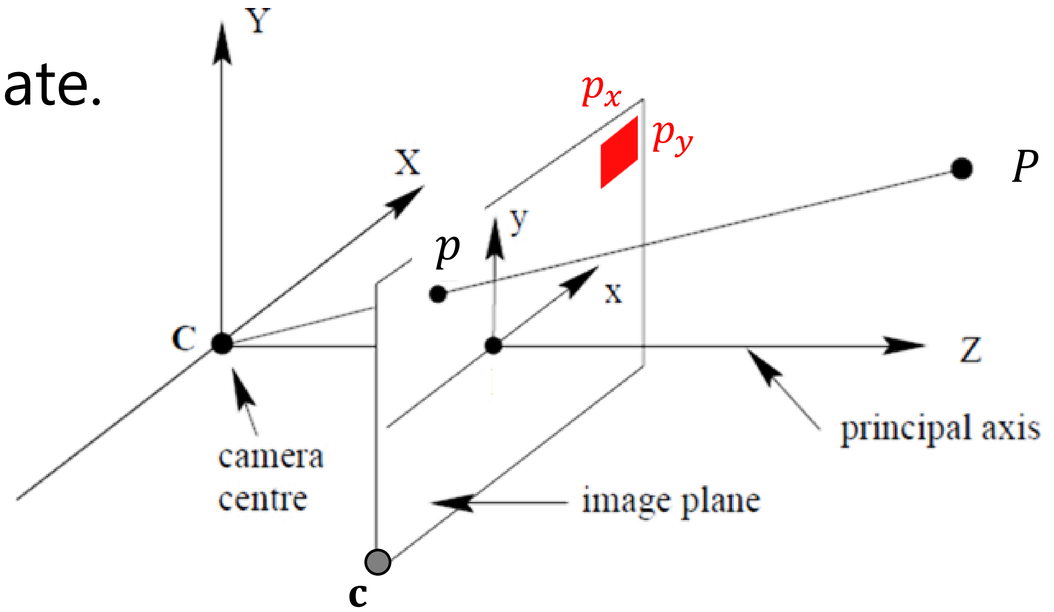
- Image plane origin has offset.
- $(x, y)^T = (fX/Z, fY/Z)^T$
- $(x', y')^T = \left(x + c_x, y + c_y\right)^T$
- $\begin{pmatrix} f & 0 & c_x \\ 0 & f & c_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} fX + c_x Z \\ fY + c_y Z \\ Z \end{pmatrix} = \begin{pmatrix} \frac{fX}{Z} + c_x \\ \frac{fY}{Z} + c_y \\ 1 \end{pmatrix} = \begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix}$

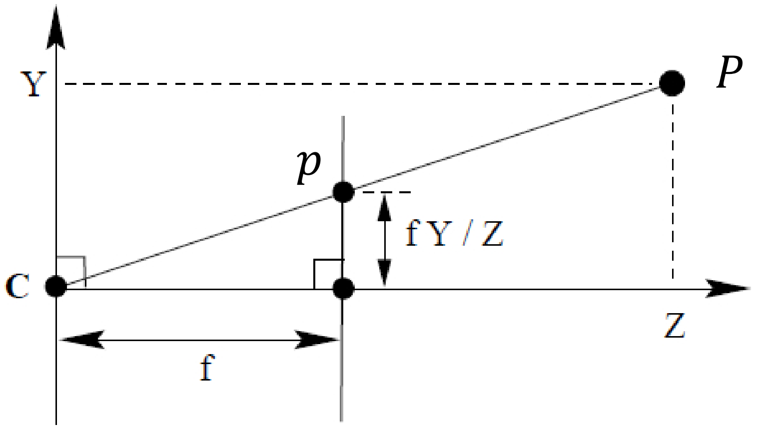




Pinhole camera model

- From spatial coordinate to pixel coordinate.
- $(x, y)^T = (fX/Z, fY/Z)^T$
- $(x', y')^T = \left(x + c_x, y + c_y\right)^T$
- $(x'', y'')^T = \left(x'/p_x, y'/p_y\right)^T$
- $\begin{pmatrix} \frac{f}{p_x} & 0 & \frac{c_x}{p_x} \\ 0 & \frac{f}{p_y} & \frac{c_y}{p_y} \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} \frac{f}{p_x} X + \frac{c_x}{p_x} Z \\ \frac{f}{p_y} Y + \frac{c_y}{p_y} Z \\ Z \end{pmatrix} = \begin{pmatrix} (\frac{fX}{Z} + c_x)/p_x \\ (\frac{fY}{Z} + c_y)/p_y \\ 1 \end{pmatrix} = \begin{pmatrix} x'' \\ y'' \\ 1 \end{pmatrix}$

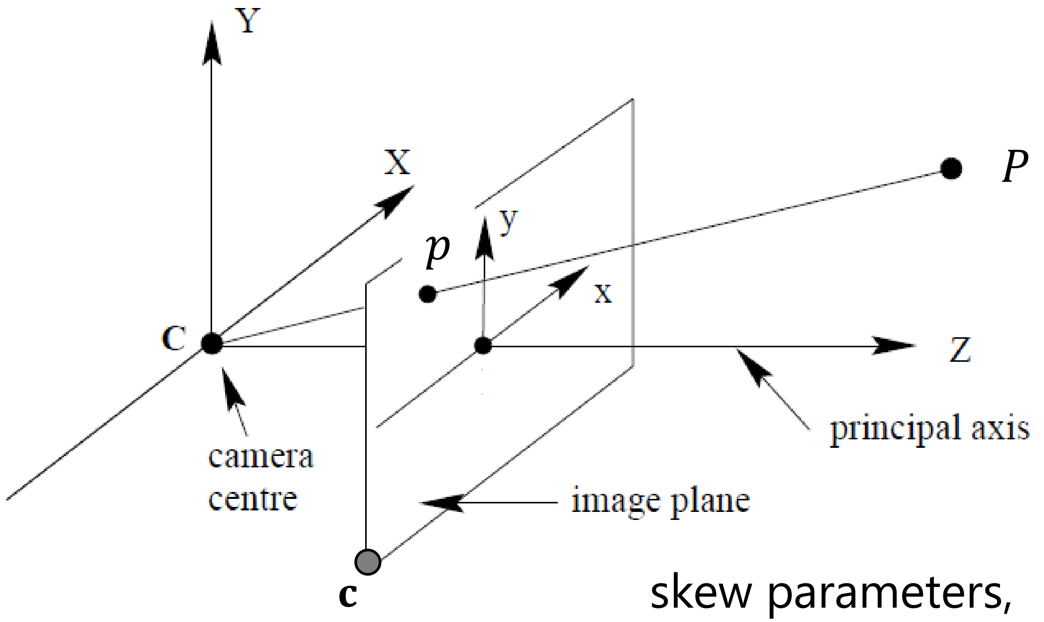




Pinhole camera model

- Redefine the parameters.
- $(x, y)^T = (fX/Z, fY/Z)^T$
- $(x', y')^T = \left(x + c_x, y + c_y\right)^T$
- $(x'', y'')^T = \left(x'/p_x, y'/p_y\right)^T$
- $\begin{pmatrix} \frac{f}{p_x} & 0 & \frac{c_x}{p_x} \\ 0 & \frac{f}{p_y} & \frac{c_y}{p_y} \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} \frac{f}{p_x} X + \frac{c_x}{p_x} Z \\ \frac{f}{p_y} Y + \frac{c_y}{p_y} Z \\ Z \end{pmatrix} = \begin{pmatrix} (\frac{fX}{Z} + c_x)/p_x \\ (\frac{fY}{Z} + c_y)/p_y \\ 1 \end{pmatrix} = \begin{pmatrix} x'' \\ y'' \\ 1 \end{pmatrix}$
- Let $f_x = f/p_x, f_y = f/p_y, c_x{}' = c_x/p_x, c_y{}' = c_y/p_y$

focal length and image origin in pixel units

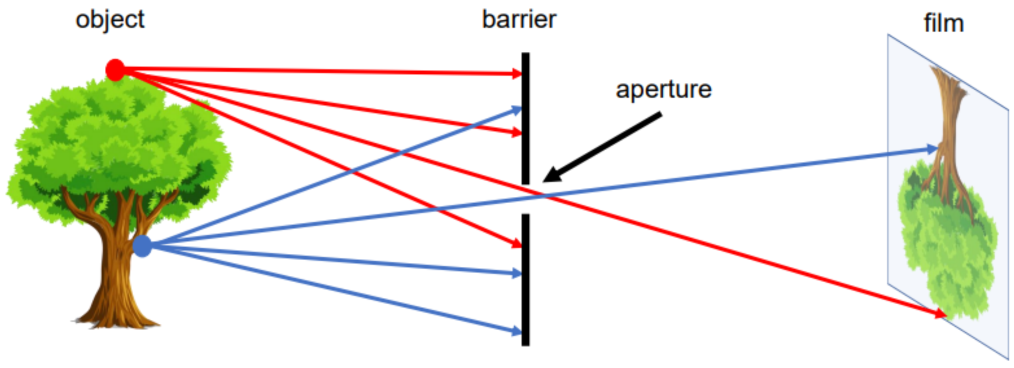


skew parameters, in most case 0

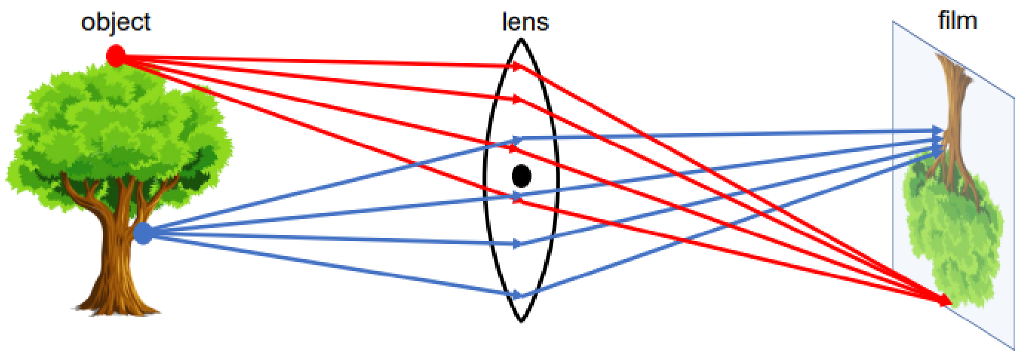
$$\underbrace{\begin{pmatrix} f_x & \alpha & c_x{}' \\ 0 & f_y & c_y{}' \\ 0 & 0 & 1 \end{pmatrix}}_{K} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} x'' \\ y'' \\ 1 \end{pmatrix}$$

Nonlinearity in camera model

- Real cameras are not pinhole. They use lens.



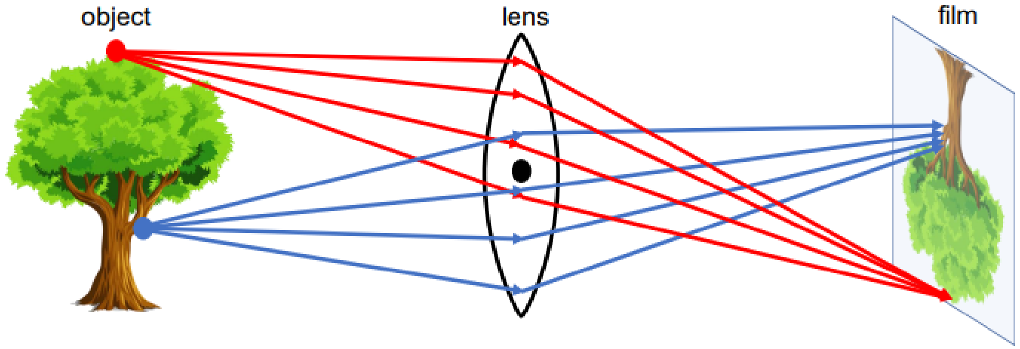
Pinhole camera

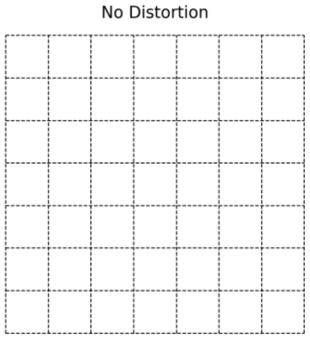


Camera with lens

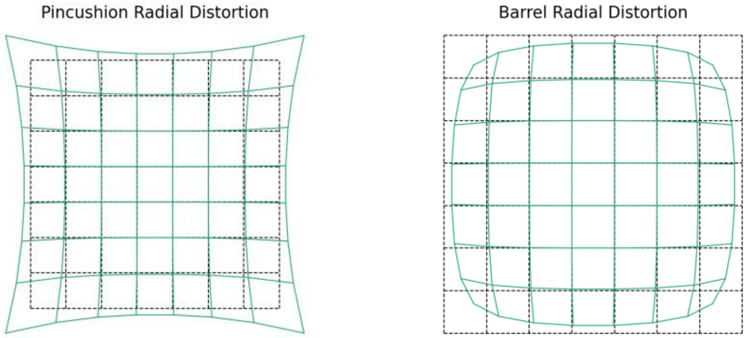
Nonlinearity in camera model

- Real cameras are not pinhole. They use lens.
- Lens imperfection causes two types of distortion: radial and tangential.



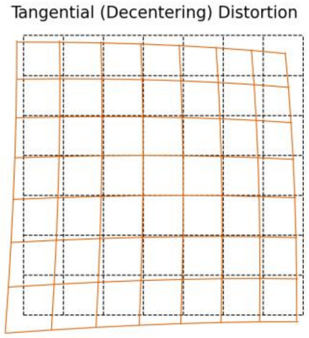


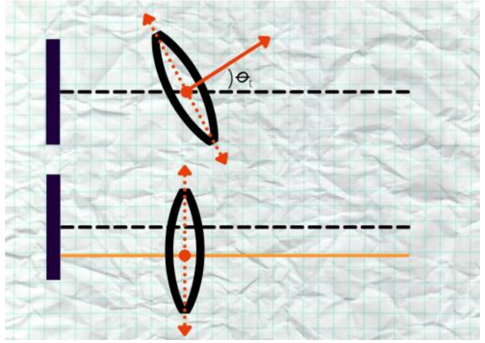
No distortion



Radial distortion

caused by imperfect lens shape



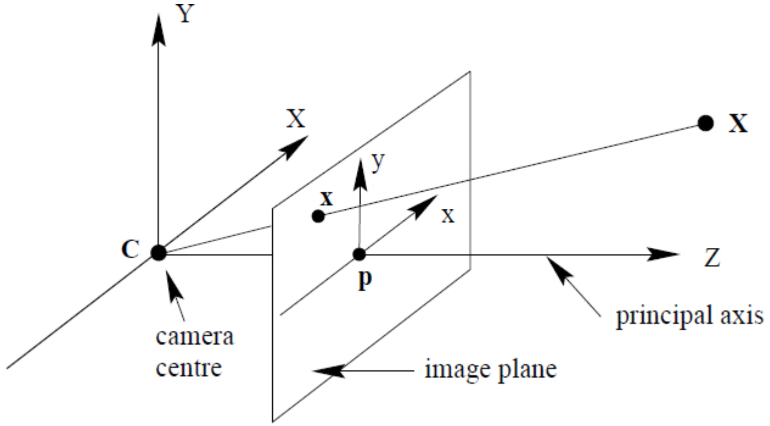


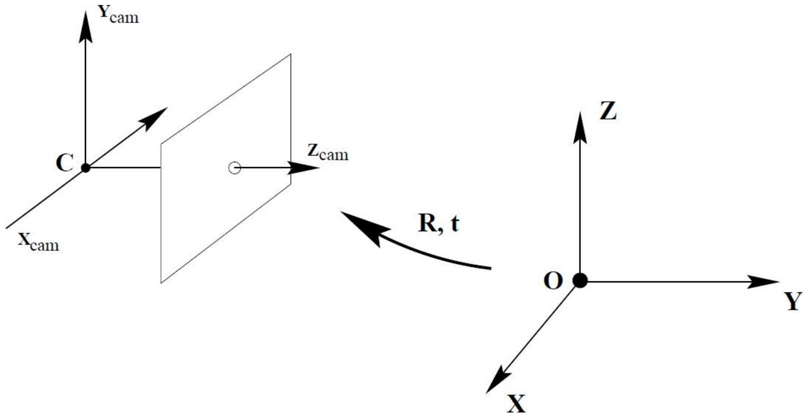
Tangential distortion

caused by imperfect lens placement

Source: Tangram Vision

Camera intrinsic and extrinsic geometry





Intrinsic

$$\begin{pmatrix} x_I \\ y_I \\ 1 \end{pmatrix} = \begin{pmatrix} f_x & \alpha & c_x{}' \\ 0 & f_y & c_y{}' \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_C \\ y_C \\ z_C \end{pmatrix}$$

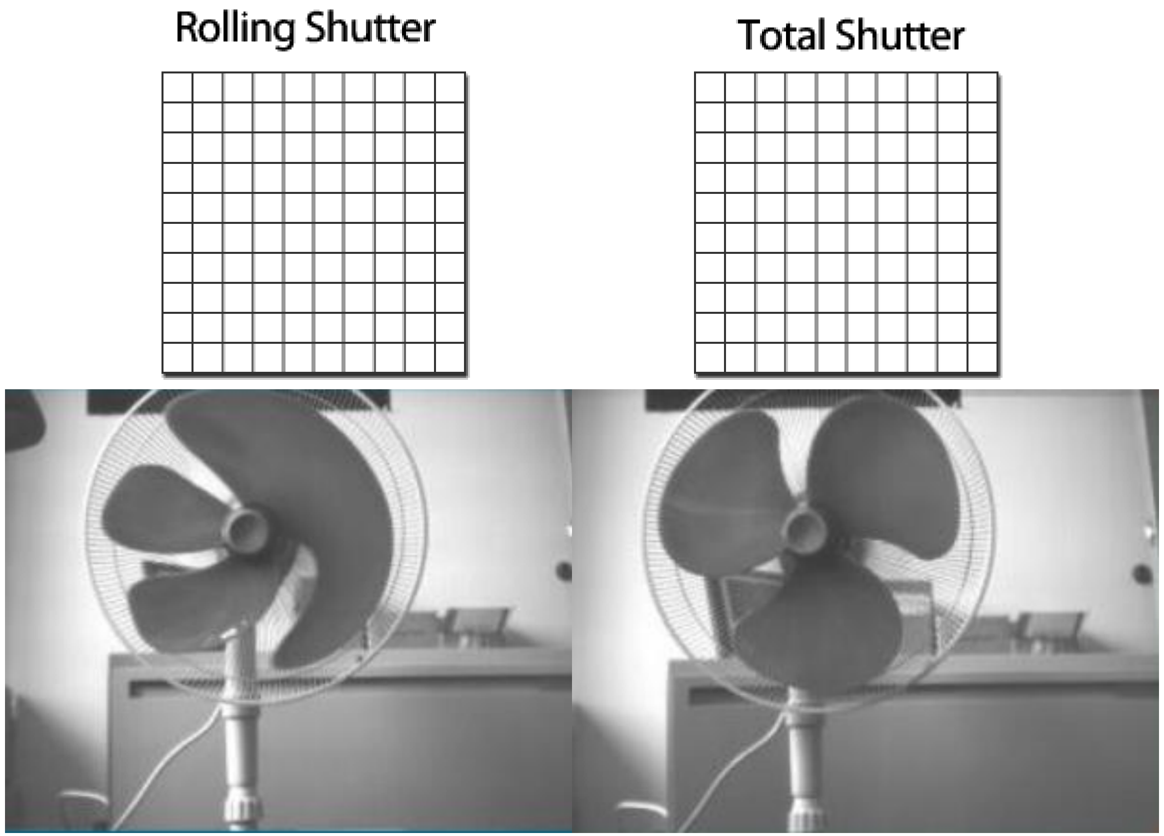
Extrinsic

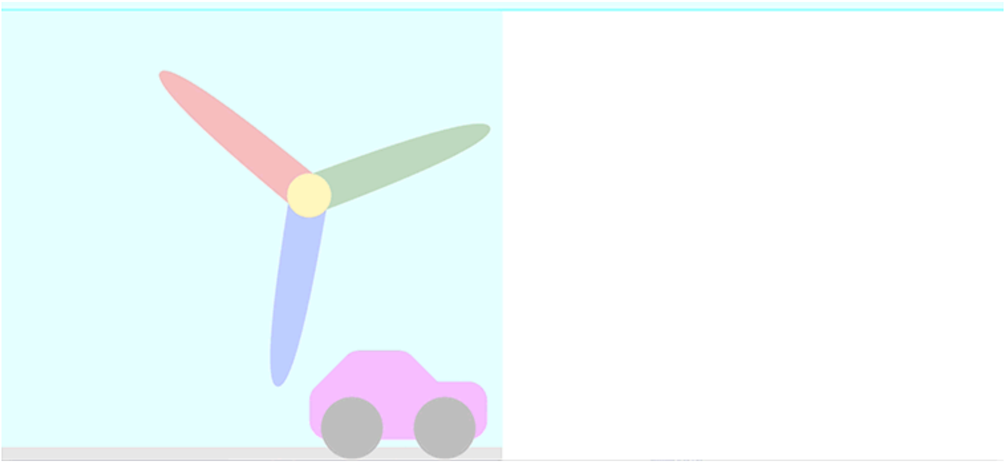
$$\begin{pmatrix} x_C \\ y_C \\ z_C \\ 1 \end{pmatrix} = \begin{pmatrix} R & t \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_W \\ y_W \\ z_W \\ 1 \end{pmatrix}$$

$$\boxed{\begin{pmatrix} x_I \\ y_I \\ 1 \end{pmatrix} = \begin{pmatrix} f_x & \alpha & c_x{}' & 0 \\ 0 & f_y & c_y{}' & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x_C \\ y_C \\ z_C \\ 1 \end{pmatrix} = (K \quad 0) \begin{pmatrix} R & t \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_W \\ y_W \\ z_W \\ 1 \end{pmatrix} = K(R \;\; t) \begin{pmatrix} x_W \\ y_W \\ z_W \\ 1 \end{pmatrix}}$$

Global shutter vs rolling shutter

- **Rolling shutter**: pixels are exposed roll by roll.
- **Global shutter**: all pixels are exposed simultaneously.





Global shutter vs rolling shutter

- Why not just use global shutter?
  - CCD sensors are being gradually replaced by CMOS sensors (due to cost, speed, power efficiency).
  - Most CCDs use global shutter. Most CMOS's use rolling shutter.

- You are likely to encounter rolling shuttle effect in video data with fast movements.

- There are CMOS sensors with global shutter.

Depth sensing

Principles for depth measurement

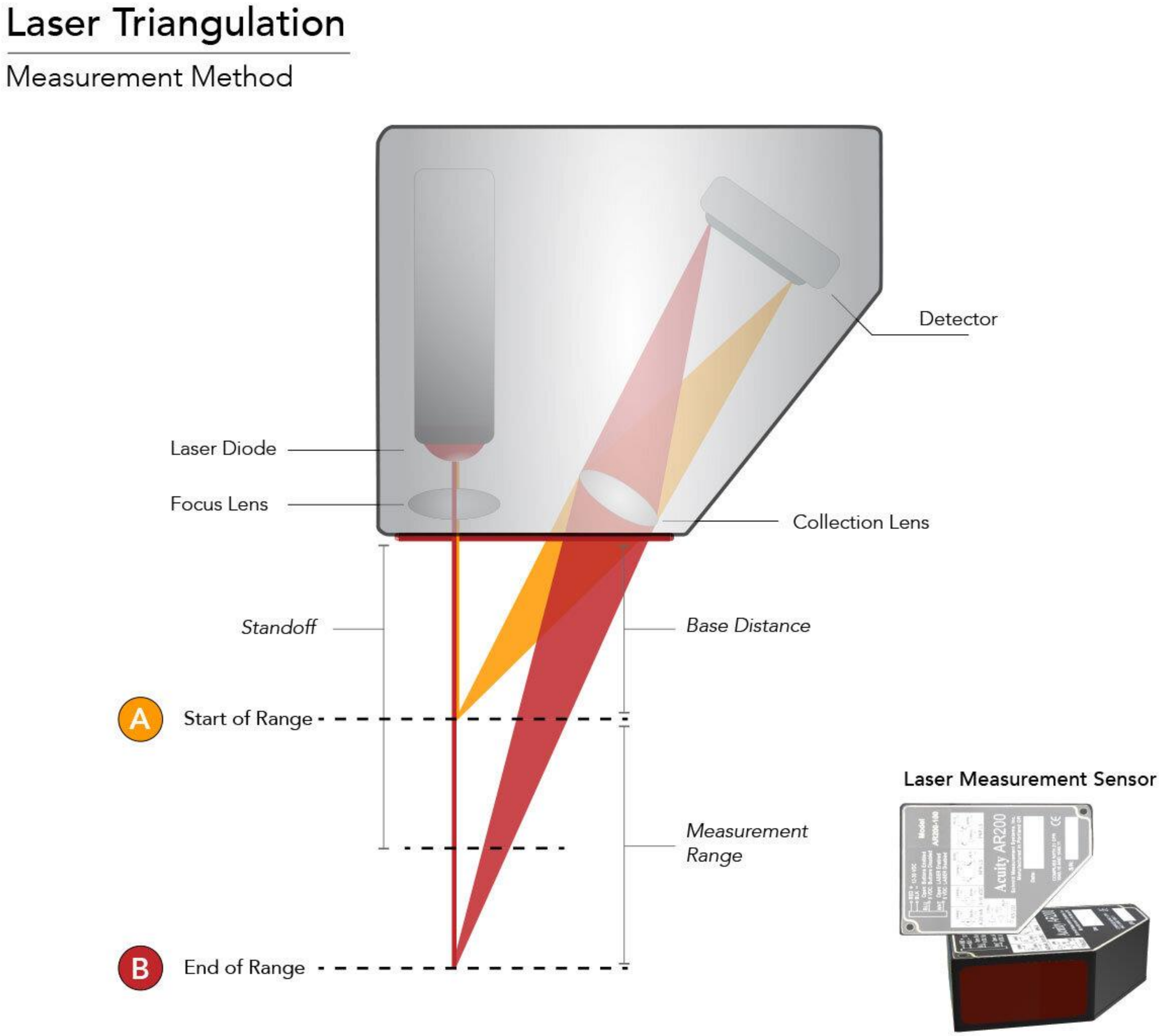
- Triangulation
  - Measuring distance from the location of the signal in the sensor.
- Time-of-Flight (ToF)
  - Measuring distance from the time for the signal to return to the sensor.

Triangulation-based depth sensing

Triangulation

Laser triangulation sensor

- The projector sends out one laser beam.
- The detector is an image sensor with one row of pixels.
- One-to-one correspondence between the range and the signal projection on the detector.



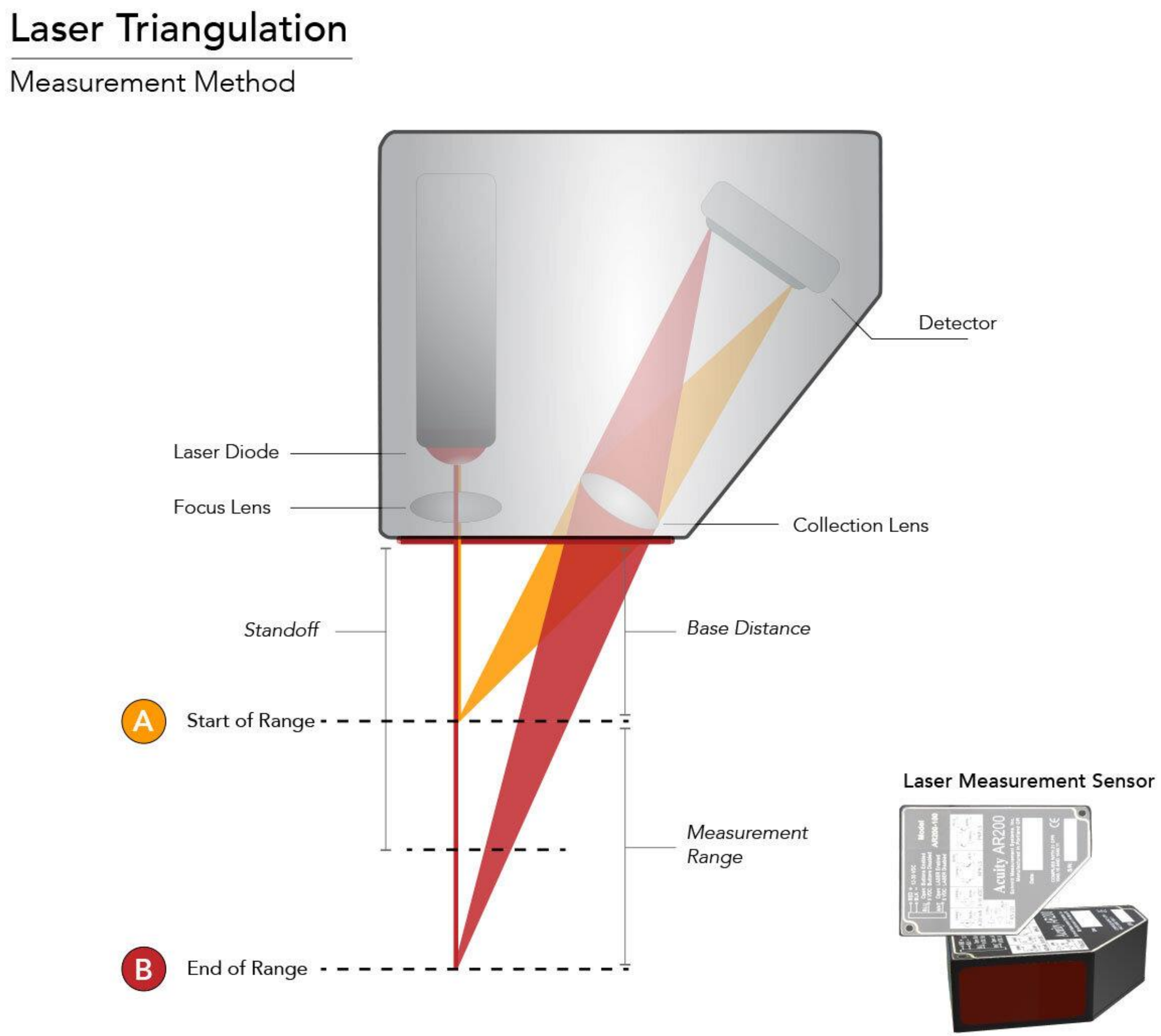
https://www.acuitylaser.com/sensor-resources/measurement-principles/

Triangulation

- Many robot vacuums use laser triangulation sensors for range measurement.
- Rotate the sensor to get range measurement of the full surrounding view.

Triangulation

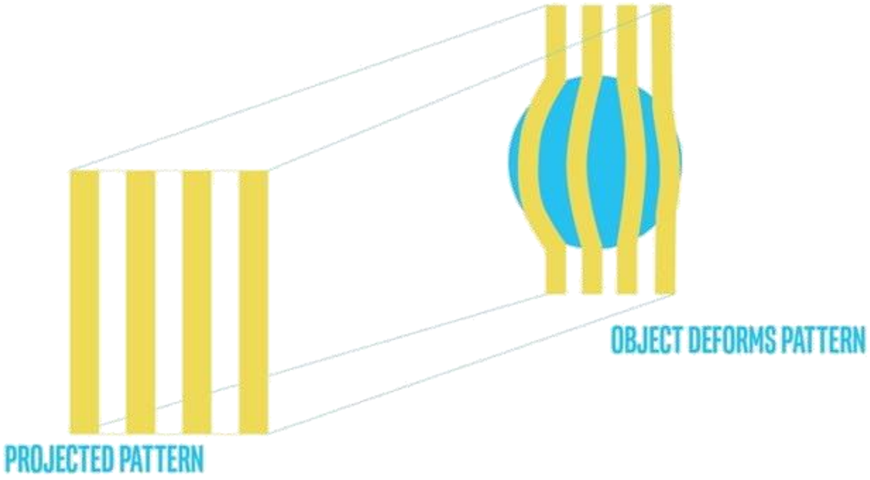
- Advantage: cheap 2D range map.
- Limitations:
  - 2D.
  - Error increases with range.
  - Sensitive to environment lighting.



Structured light

Generalize laser triangulation sensors to 2D:

- Cast a known pattern of light (instead of a single beam) to the environment.
- Detect the light pattern using an image sensor.
- The displacement of the pattern reveals the depth information.

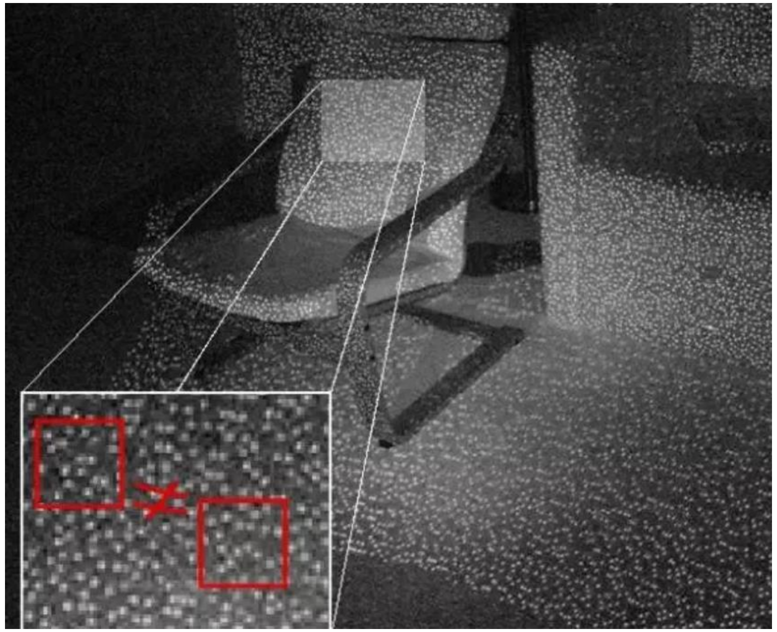


https://www.intelrealsense.com/beginners-guide-to-depth/

Structured light

Generalize laser triangulation sensors to 2D:

- How to distinguish different light beams?
- The local patterns of the projected light are unique.

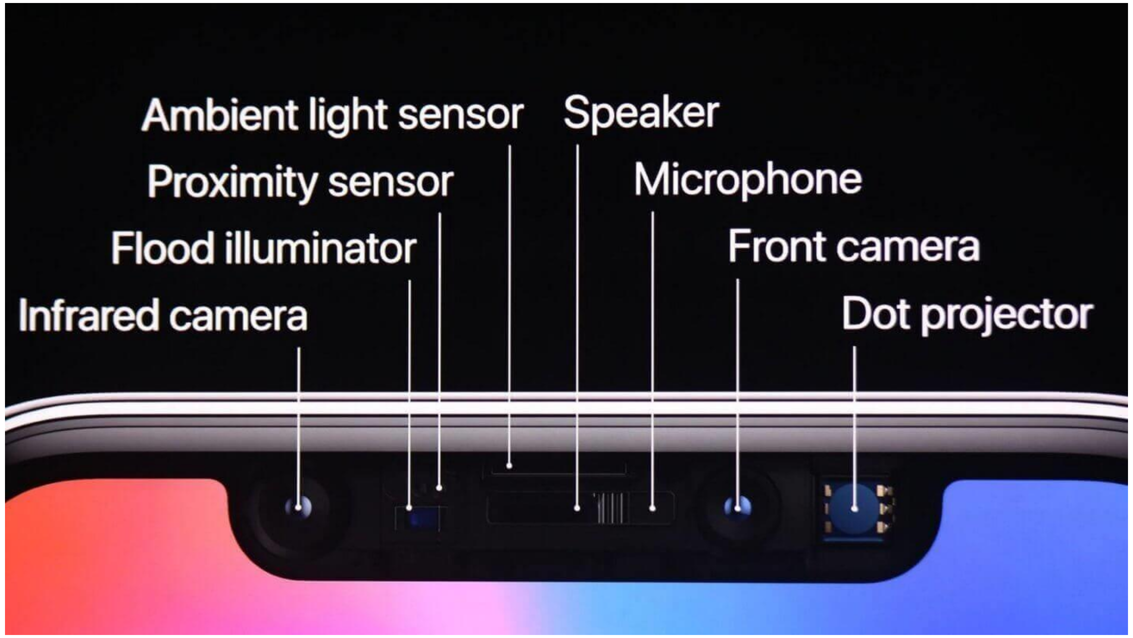


https://zhuanlan.zhihu.com/p/32200287

Structured light

- Applications

Kinect V1



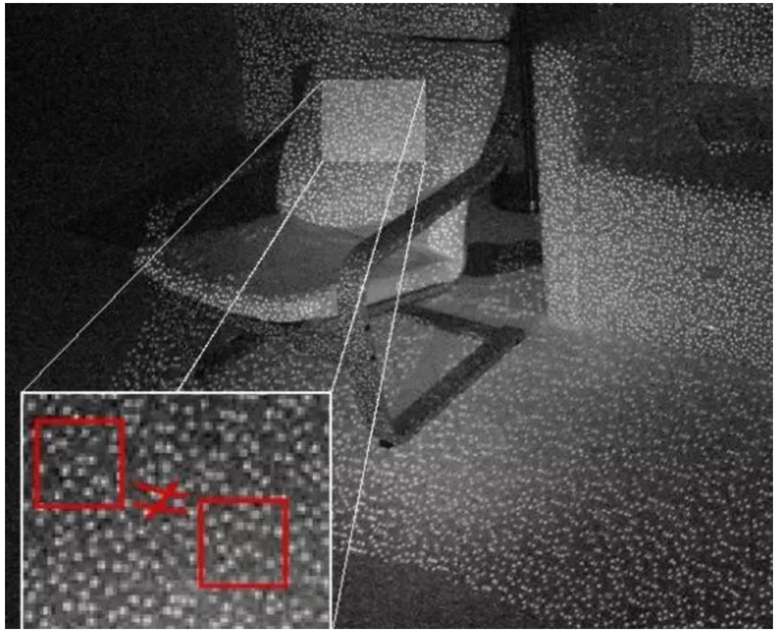
iPhone Face ID

https://zhuanlan.zhihu.com/p/32200287

https://all3dp.com/iphone-x-mean-3d-scanning/

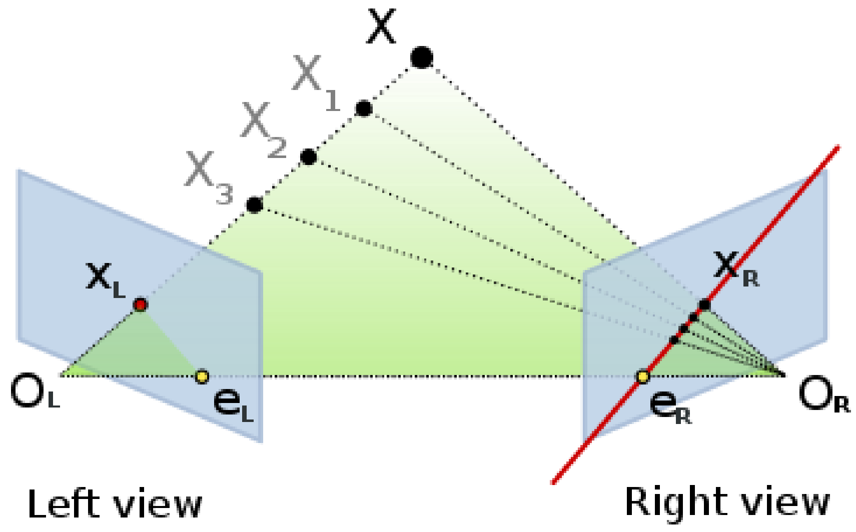
Structured light

- Advantages:
  - Works in low-light and texture-less environment.
  - Good accuracy.
  - Good resolution.
- Limitations:
  - Less reliable in strong-light conditions.
  - Error increases with range.



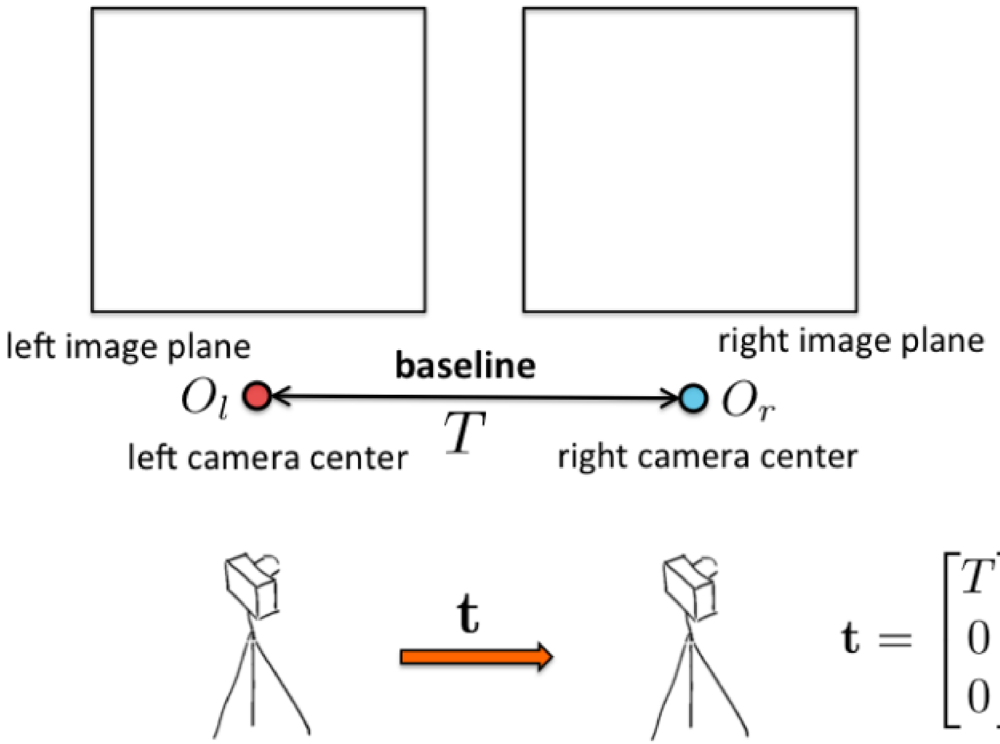
Stereo camera

- Projector + Camera → Camera + Camera
  - A point in the left view (a hypothetical beam of light) corresponds to a line in the right view.
  - One-to-one correspondence between the depth and the projected location in the right view.
- Questions:
  - How to find corresponding X_L and X_R?
  - How to calculate depth from X_L and X_R?



Stereo camera

- Consider a special case: the two cameras are parallel.
  - i.e. the right camera is just some distance to the right of left camera.



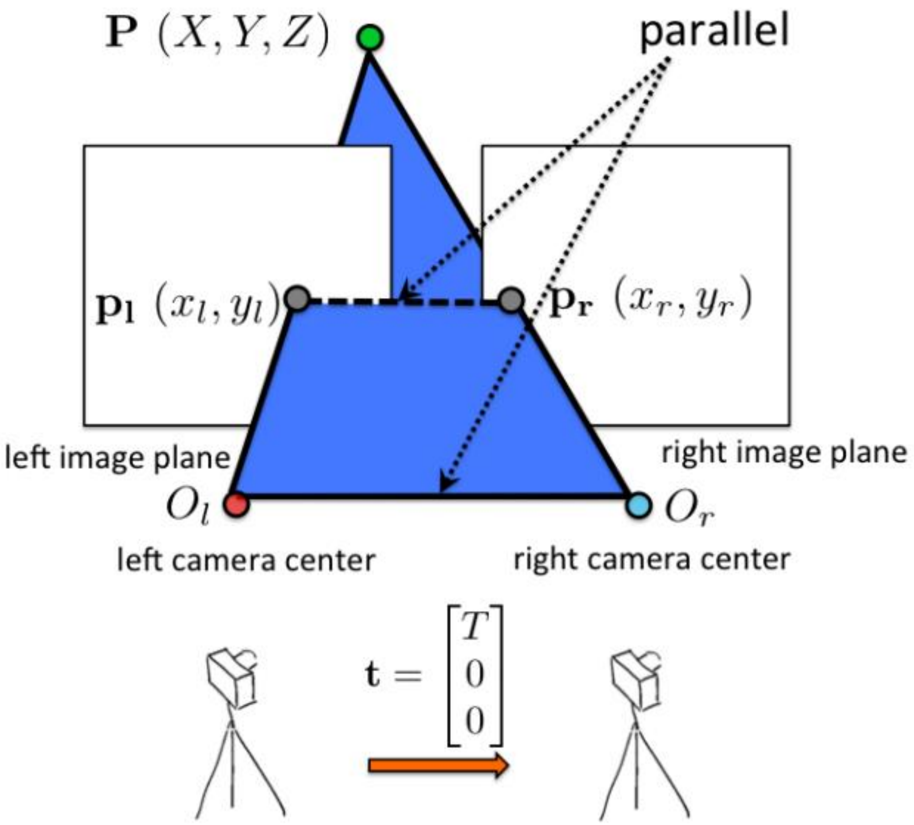
$$\mathbf{t} = \begin{bmatrix} T \\ 0 \\ 0 \end{bmatrix}$$

The right camera is shifted to the right in X direction

https://www.cs.toronto.edu/~fidler/slides/2015/CSC420/lecture12_hres.pdf

Stereo camera

- Consider a special case: the two cameras are parallel.



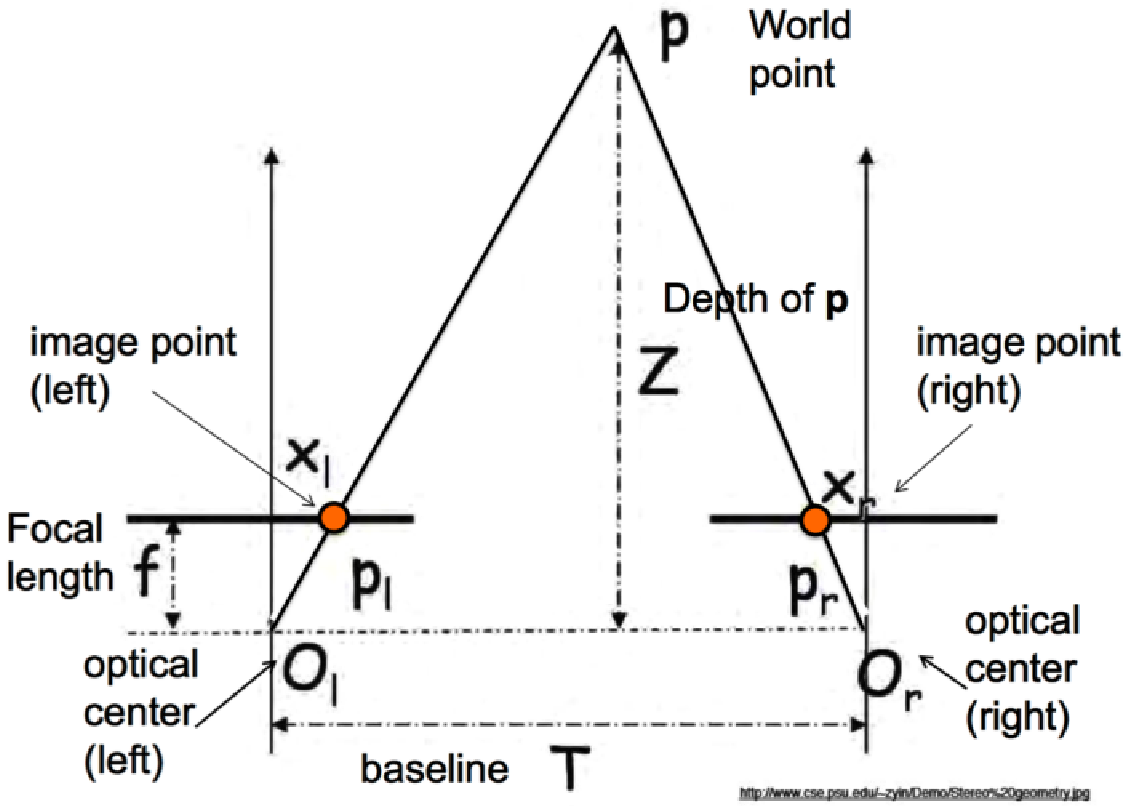
So: $y_r = y_l$

Only need to search for matching points on a horizontal line!

https://www.cs.toronto.edu/~fidler/slides/2015/CSC420/lecture12_hres.pdf

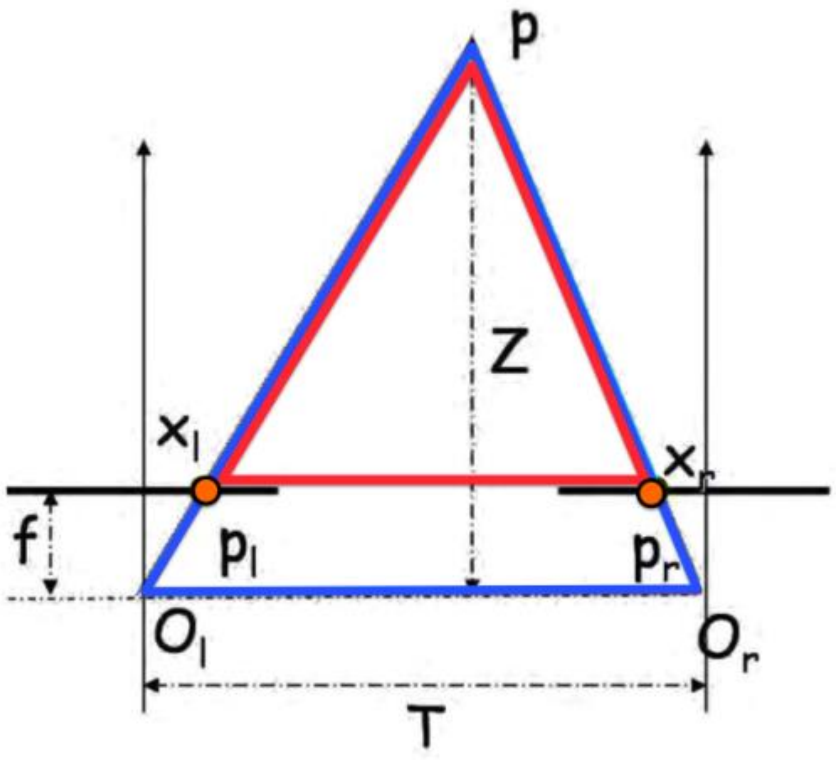
Stereo camera

- Since our points $\mathbf{p_l}$ and $\mathbf{p_r}$ lie on a horizontal line, we can forget about y_l for a moment (it doesn't seem important). Let's look at the camera situation from the birdseye perspective instead. Let's see if we can find a connection between x_l, x_r and Z (because Z is what we want).



Stereo camera

- We can then use similar triangles to compute the depth of the point P



Similar triangles:

$$\frac{T}{Z} = \frac{T + x_r - x_l}{Z - f}$$

$$\boxed{Z = \frac{f \cdot T}{x_l - x_r}}$$

baseline

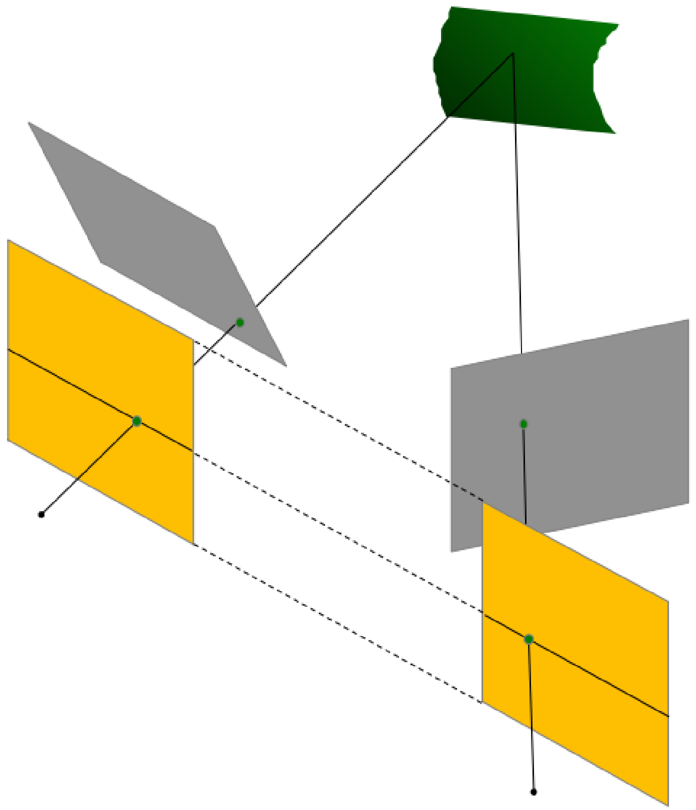
focal length

disparity

Effective range dominated by the baseline.

Stereo camera

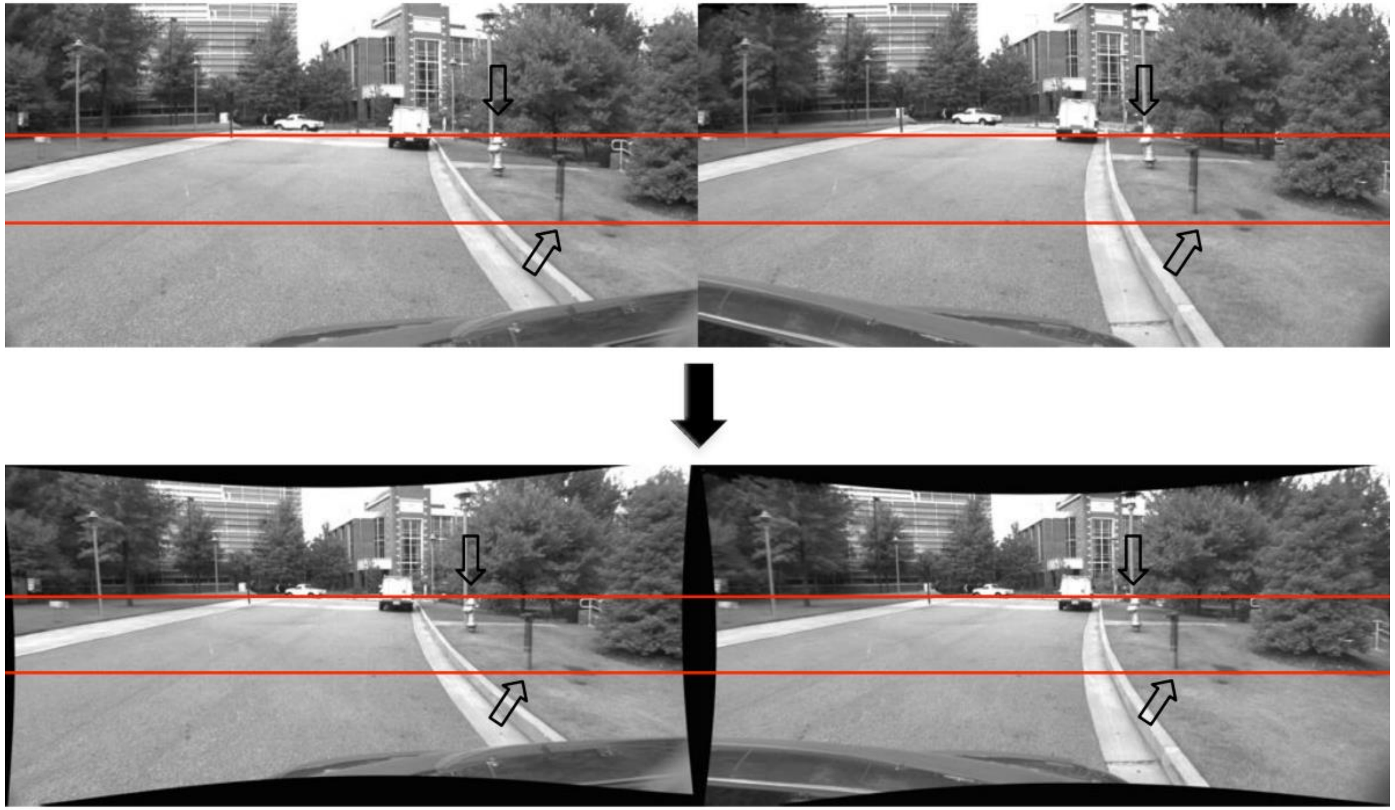
- What if the cameras are not parallel?
- We can apply **image rectification** to synthesize parallel cameras.
- Apply two homography matrices (3x3) to transform the two images respectively.
- Details skipped in this lecture.
- Some datasets did the rectification for you.



https://www.cs.cmu.edu/~16385/s17/Slides/13.1_Stereo_Rectification.pdf
https://www.cs.toronto.edu/~fidler/slides/2015/CSC420/lecture13_hres.pdf

Stereo camera

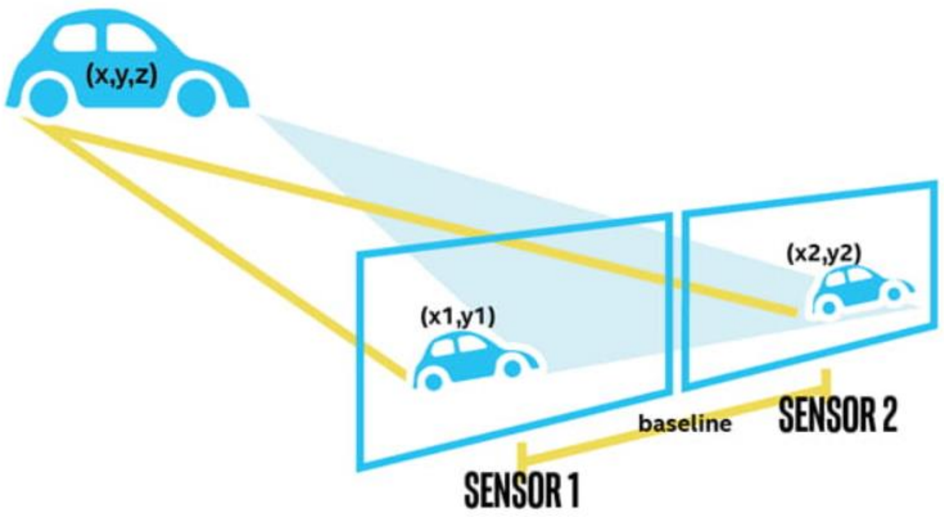
- After camera undistortion and rectification, matches can be found on a horizontal line.



Slides:C. Beall

Stereo camera

- Advantages:
  - Low cost. No special hardware other than cameras.
  - Works both indoor and outdoor.
- Limitations:
  - Fails in dark and texture-less environment
  - Error increases with range.
  - Requires good matching algorithms.



https://www.intelrealsense.com/beginners-guide-to-depth/

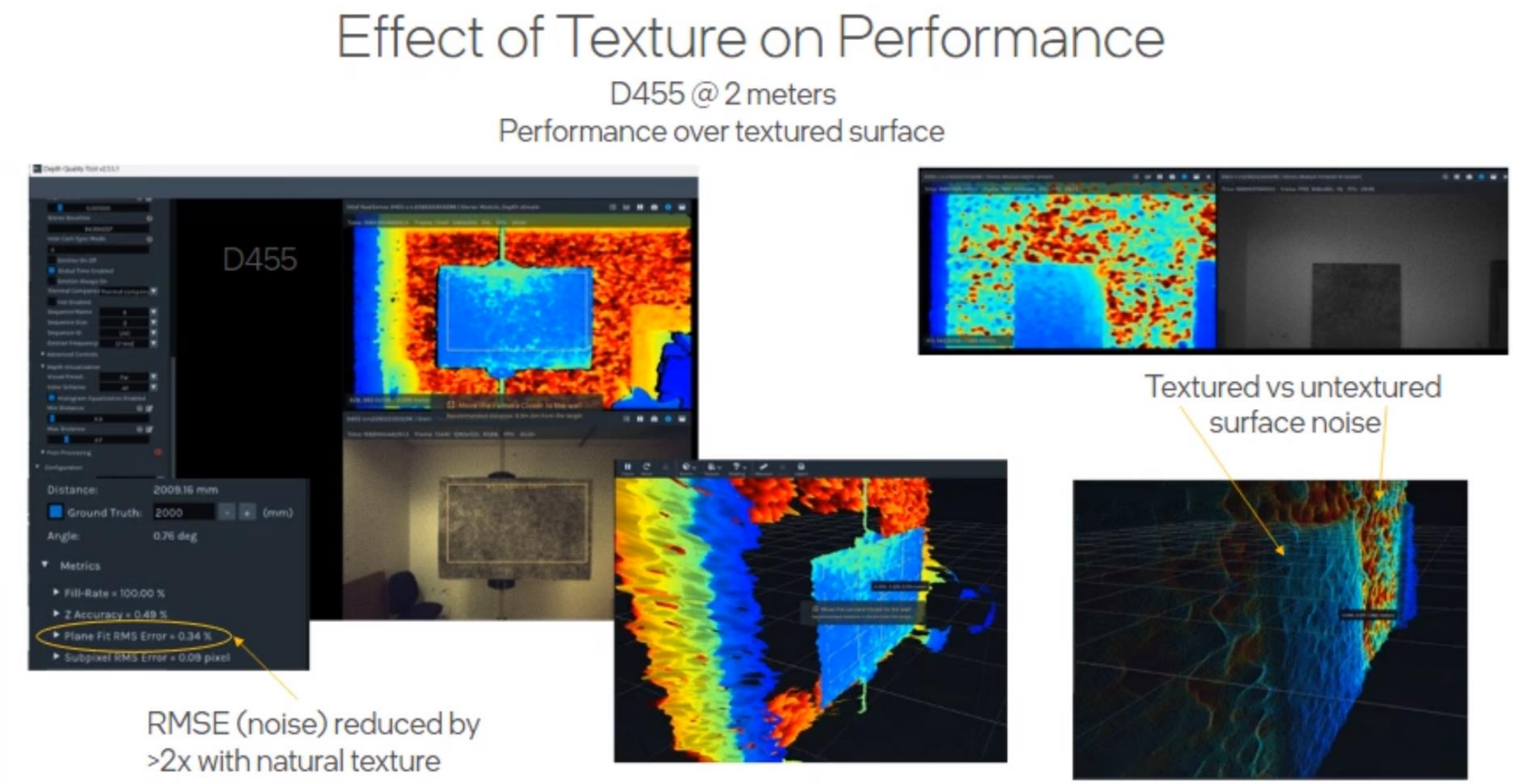
Active stereo camera

- Combine stereo camera with the idea of structured light.

Intel® RealSense™ Depth Camera D455

Active stereo camera

- Combine stereo camera with the idea of structured light.



https://webinar.intel.com/Unlocking-new-dimensions-in-stereo-depth-twice-the-range-reg

ToF-based depth sensing

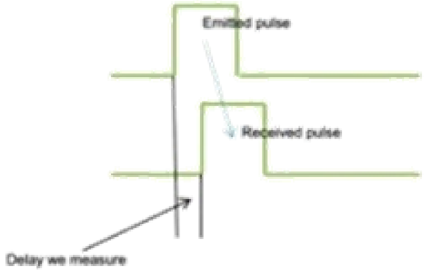
Time-of-Flight depth cameras

- Measuring distance from the time for the emitted light signal to hit the scene and return back to the sensor.
- Compared with triangulation:
  - Advantage of ToF:
    - Longer range.
    - Works in low-light and texture-less environment.
  - Limitations:
    - Less reliable in strong-light environment.

Time-of-Flight depth cameras

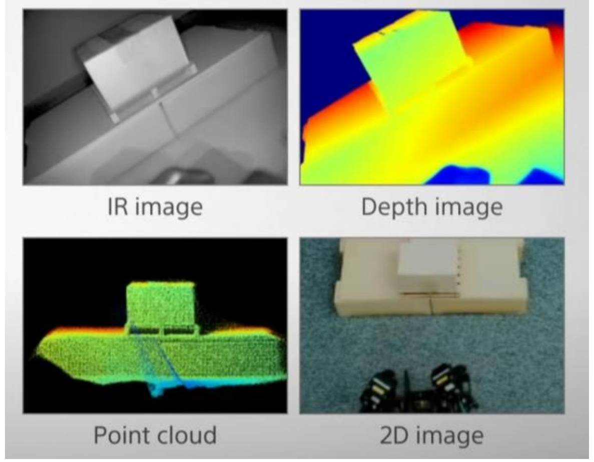
- Measuring distance from the time for the emitted light signal to hit the scene and return back to the sensor.
- Indirect ToF (**iToF**)
  - Measures *<u>phase shift</u>* of modulated continuous light signal.
  - Measure the entire scene in a single scan.

- Direct ToF (**dToF**)
  - Measures *<u>transit time</u>* of reflected pulse of light from an object.
  - Measure the distance per beam of light.



Time-of-Flight depth cameras

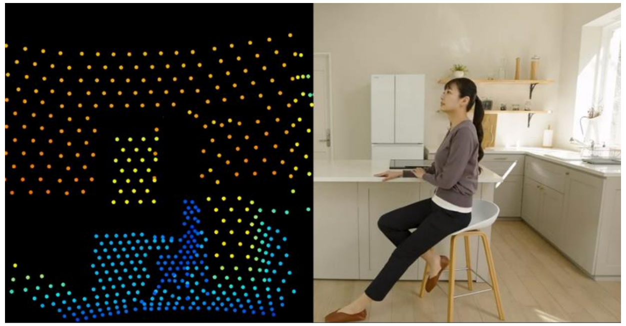
Indirect ToF (iToF)



- Advantages
  - Cheaper.
  - High resolution.
  - High frame rate.
- Limitations:
  - Noisier depth.
  - Less reliable in strong-light conditions.
  - Range-accuracy tradeoff.

https://www.sony-semicon.com/en/technology/industry/tof.html

Direct ToF (dToF)



- Advantages
  - High accuracy regardless of distance.
  - More robust to lighting conditions.
  - Lower power consumption.
- Limitations:
  - Higher cost.
  - Lower resolution.

https://www.sony-semicon.com/en/news/2023/2023030601.html

Time-of-Flight depth cameras

- Examples

Kinect V2 (iToF)

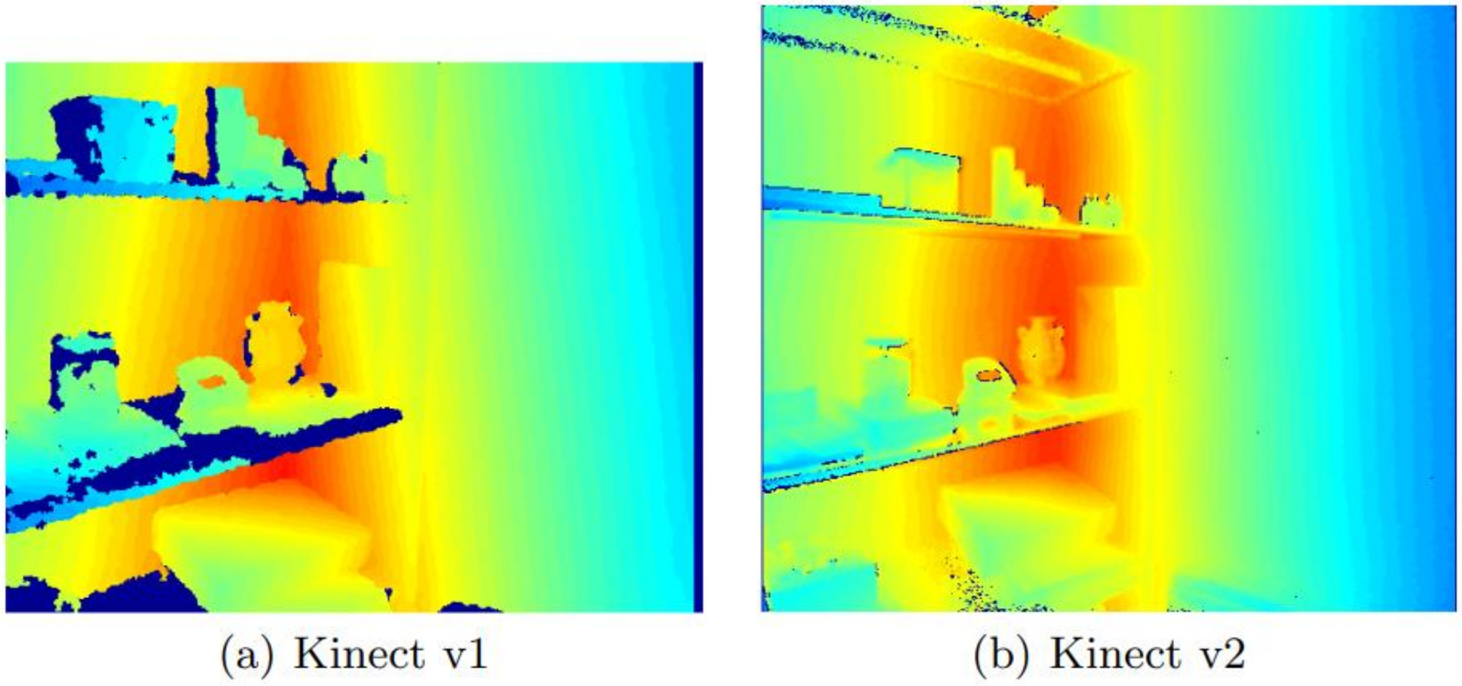
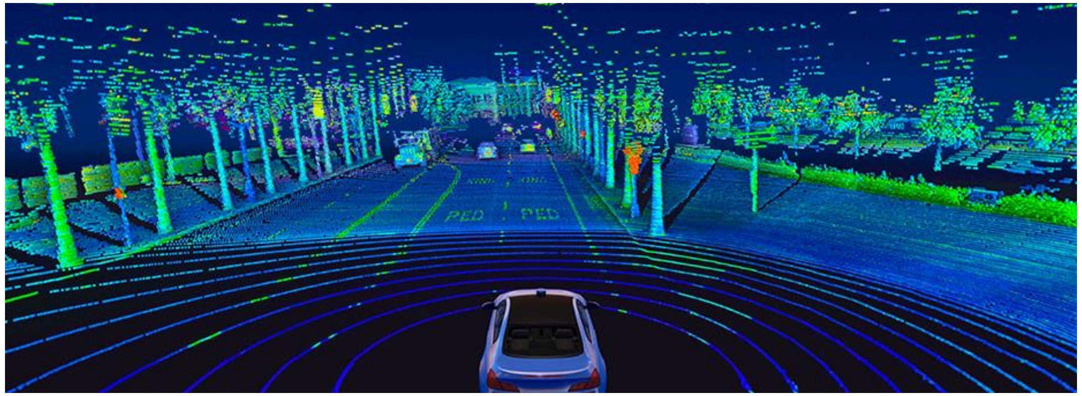


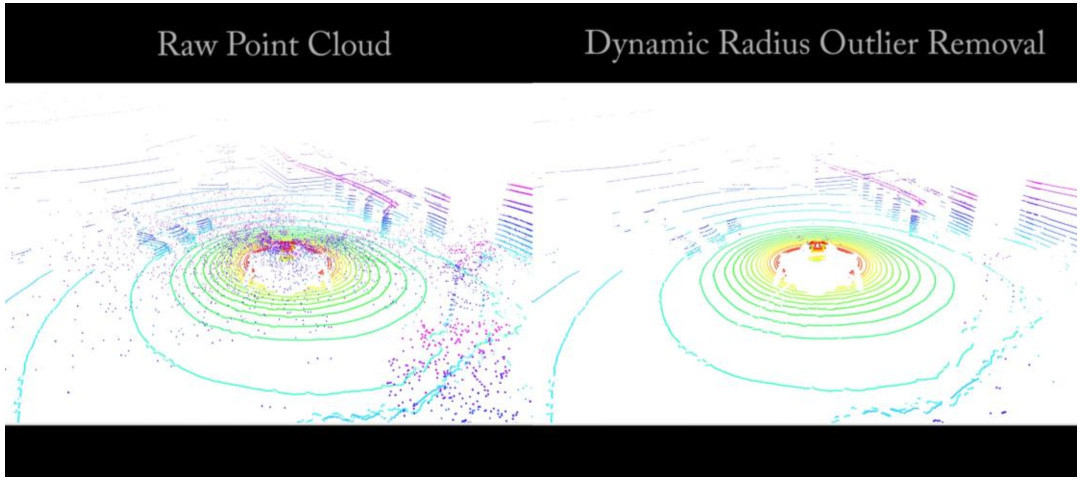
Fig. 2. Captured depth images of the same scene for the Kinect v1 and Kinect v2.

Wasenmüller, Oliver, and Didier Stricker. "Comparison of kinect v1 and v2 depth images in terms of accuracy and precision." *Computer Vision–ACCV 2016 Workshops: ACCV 2016 International Workshops, Taipei, Taiwan, November 20-24, 2016, Revised Selected Papers, Part II 13*. Springer International Publishing, 2017.

Lidar (light detection and ranging)

- Same principle as dToF, but larger power emitter, laser pulses $\lambda \approx 900{\sim}1500\ \mathrm{nm}$.
  - Longer range and work outdoor.
- Strengths:
  - Accurate range measurement (~1cm).
  - High angular resolution ($0.1^\circ{\sim}1^\circ$).
- Weaknesses:
  - Sensitive to adverse weather.
  - Expensive (~$1000).
- Applications:
  - Outdoor 3D detection and mapping.



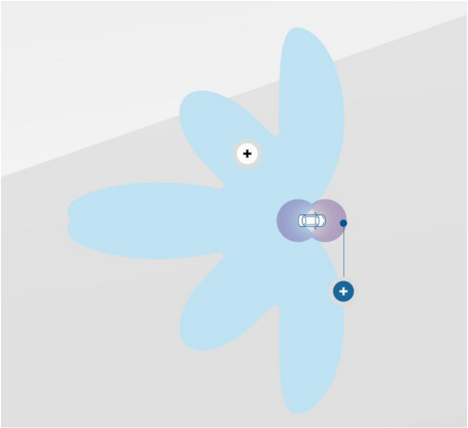


Snow brings noise in Lidar scan.

Source: Bosch

Radar (radio detection and ranging)

- Emitting and receiving millimeter waves ($\lambda \approx 1\sim10$ mm).
- Strengths:
  - Robust to rain, snow, and fog.
  - Long range (~300m).
  - Low cost (~$100).
- Weaknesses:
  - Low angular resolution ($1° \sim 10°$).
- Can be used in outdoor scenarios with emphasis on weather robustness.

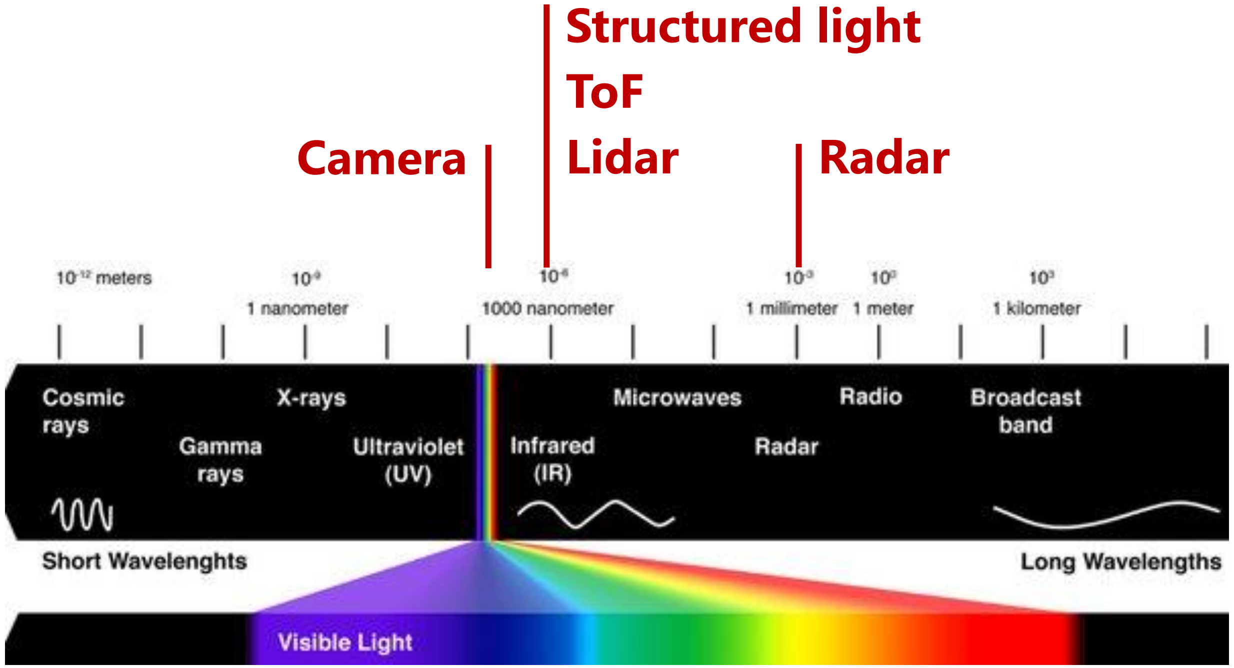


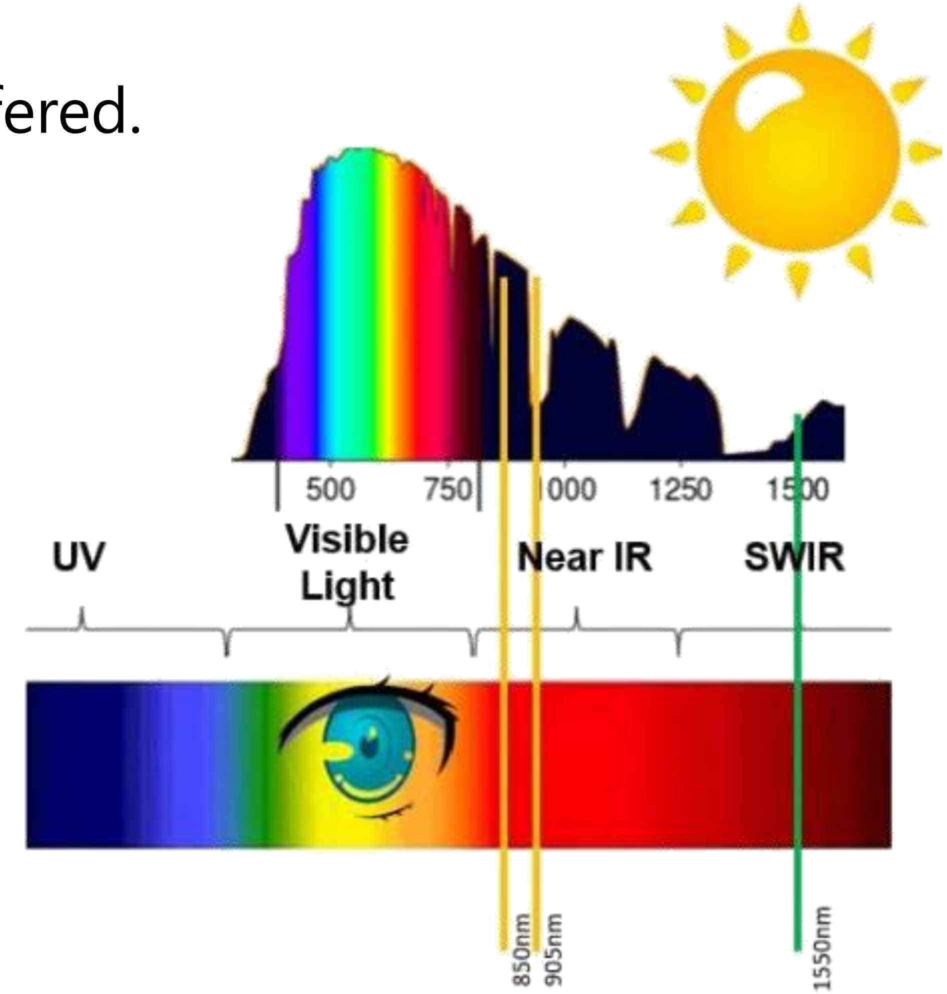
Radar and Ultrasonic coverage for a car.

* Ultrasonic sensor: ~5m range.

Spectrum of the sensors

- The shorter the wavelength, the easier to be scattered.
- The closer to visible light, the easier to be interfered.





Comparison

Property	Structured Light	Stereo Vision	LiDAR	dToF	iToF
Principle	Observes distortions in projected pattern	Compares features in two stereo images	Measures transit time of reflected light from an object	Measures transit time of reflected light from an object	Measures phase shift of modulated light pulses
Software Complexity	Very high	High	Low	Low	Medium
Relative Cost	High	Low	Varies	Low	Medium
Accuracy	µm - mm	cm	Depends on range	mm-cm	mm-cm
Operating Range	Low	~6m	Very scalable	Scalable	scalable
Low Light	Good	Weak	Good	Good	Good
Outdoor	Weak	Good	Good	Fair	Fair
Scan Speed	Slow	Medium	Slow	Fast	Very Fast
Compactness	Medium	Low	Low	High	Medium
Power Consumption	High	Low	High	Low - Medium	Medium

https://www.framos.com/en/articles/what-are-depth-sensing-cameras-and-how-do-they-work