NUS
National University of Singapore

EE5060
Smart Sensors and Instrumentation

EE5111
Industrial Control and Instrumentation

Jiang Rui, Ph.D., BEng
rui.j@nus.edu.sg
Adjunct Lecturer
Department of Electrical and Computer Engineering

Semester 1, AY2021/2022

Outline

Vision and Localization in Robotics and Autonomous Systems

- CMOS sensors and machine vision systems
- Positioning systems
- Vision-based pose estimation

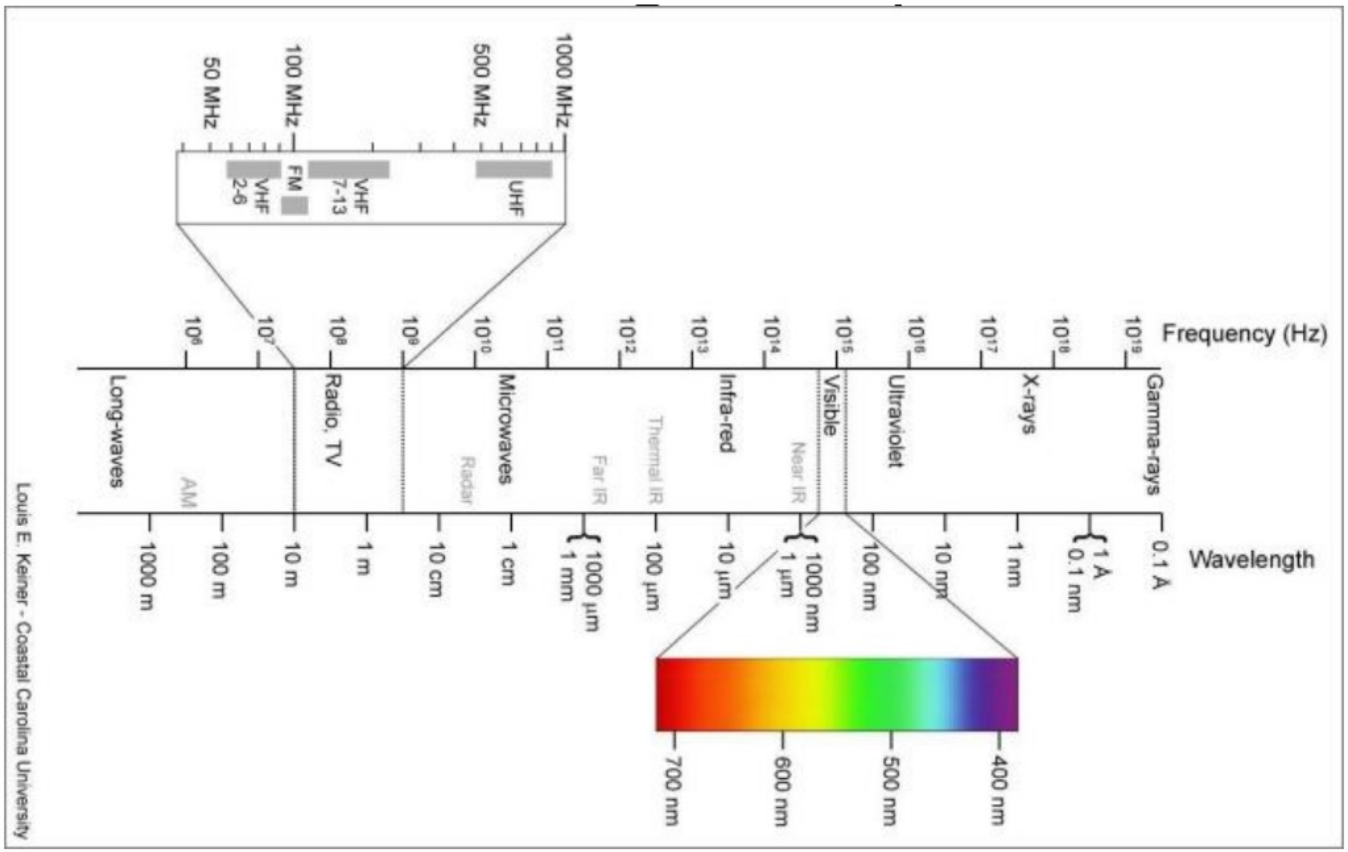
Learning Objectives

- In this lecture, you are expected to:

1. Explain the basic working principles and characteristics of CMOS Image Sensors (CIS);
2. Understand the principles of commonly-used positioning sensors and methods in robotics and autonomous systems.
3. Understand the principles of vision-based localization / pose estimation.

Light Basics

- Electromagnetic Spectrum
  - Light is a very narrow band in the electromagnetic spectrum

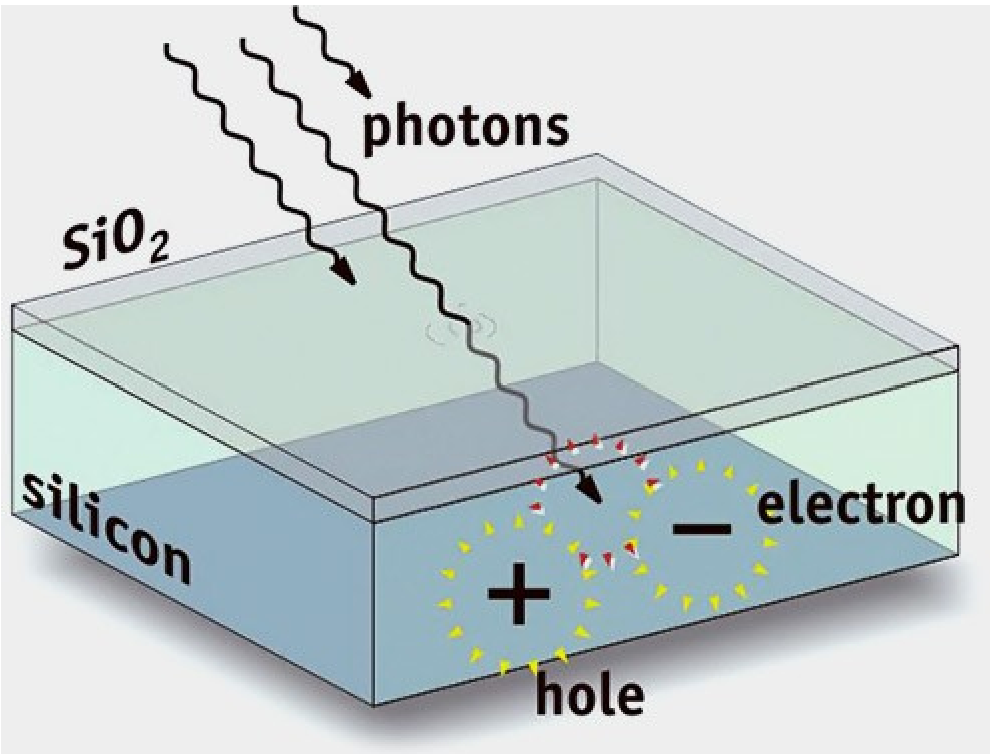


Light Basics

- We are primarily interested in wavelengths from 200 - 1100 nm
  - Visible Light: 400 – 750 nm
  - Near Ultra Violet: 200 – 400 nm
  - **Near Infrared Region: 750 – 1100 nm**

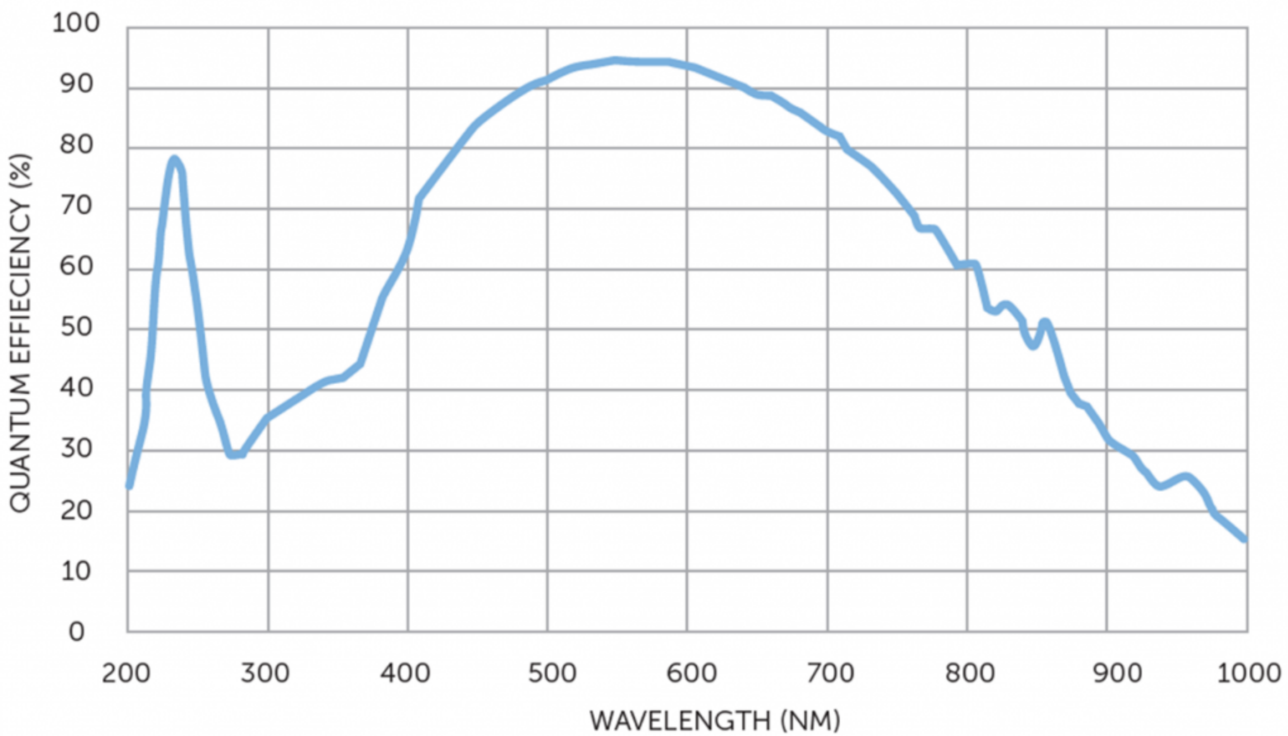
Photoelectric Effect: Photons to Electrons

- Light photons hitting a Silicon surface will dislodge electrons
- Number of electrons released depends on intensity and wavelength



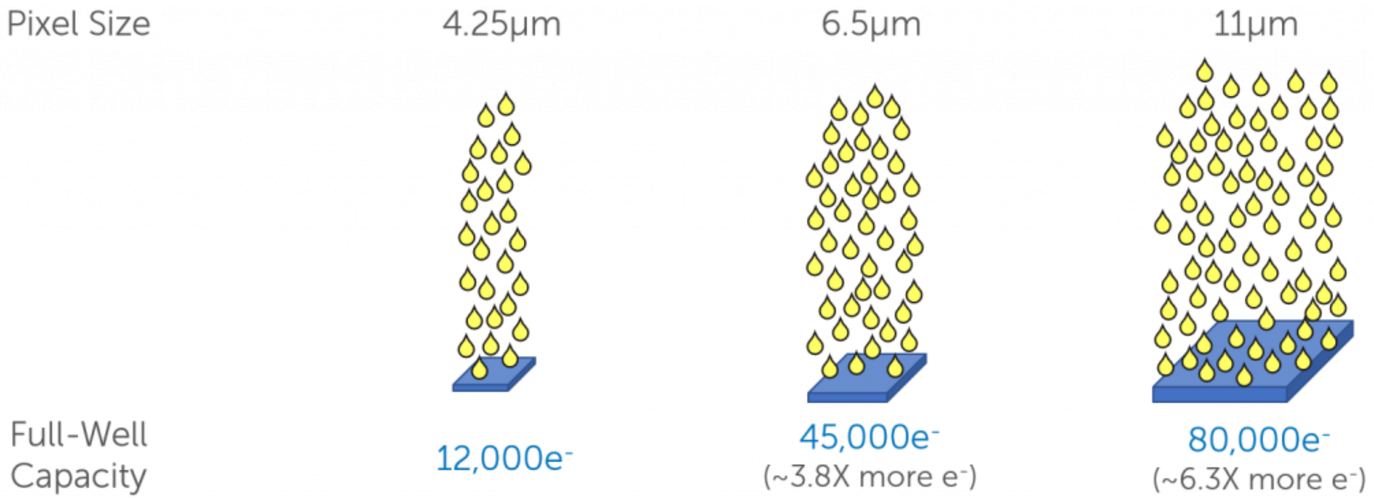
Quantum Efficiency (QE)

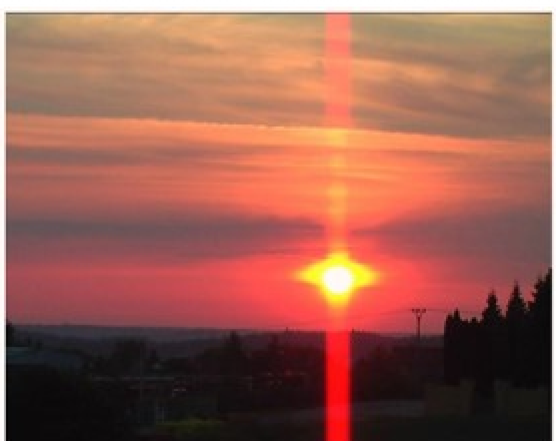
- The ratio of **light** that the sensor converts into **charge**
  - 60% QE = For every 10 photons hitting a pixel, 6 electrons are released.
  - QE response is sensor specific. Camera design does not affect the QE curve.

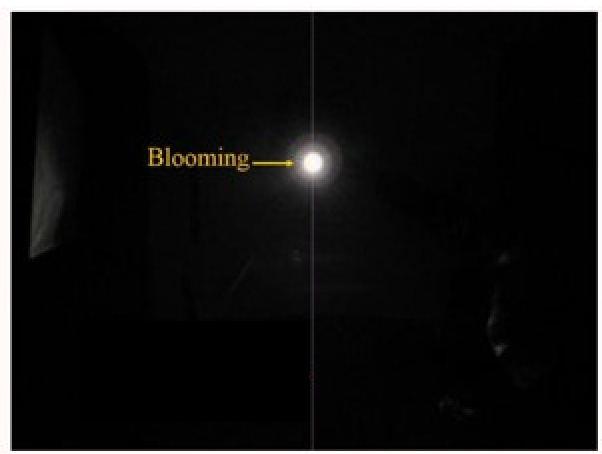


Full Well Capacity

- The amount of charge that can be stored within an individual pixel without the pixel becoming saturated.

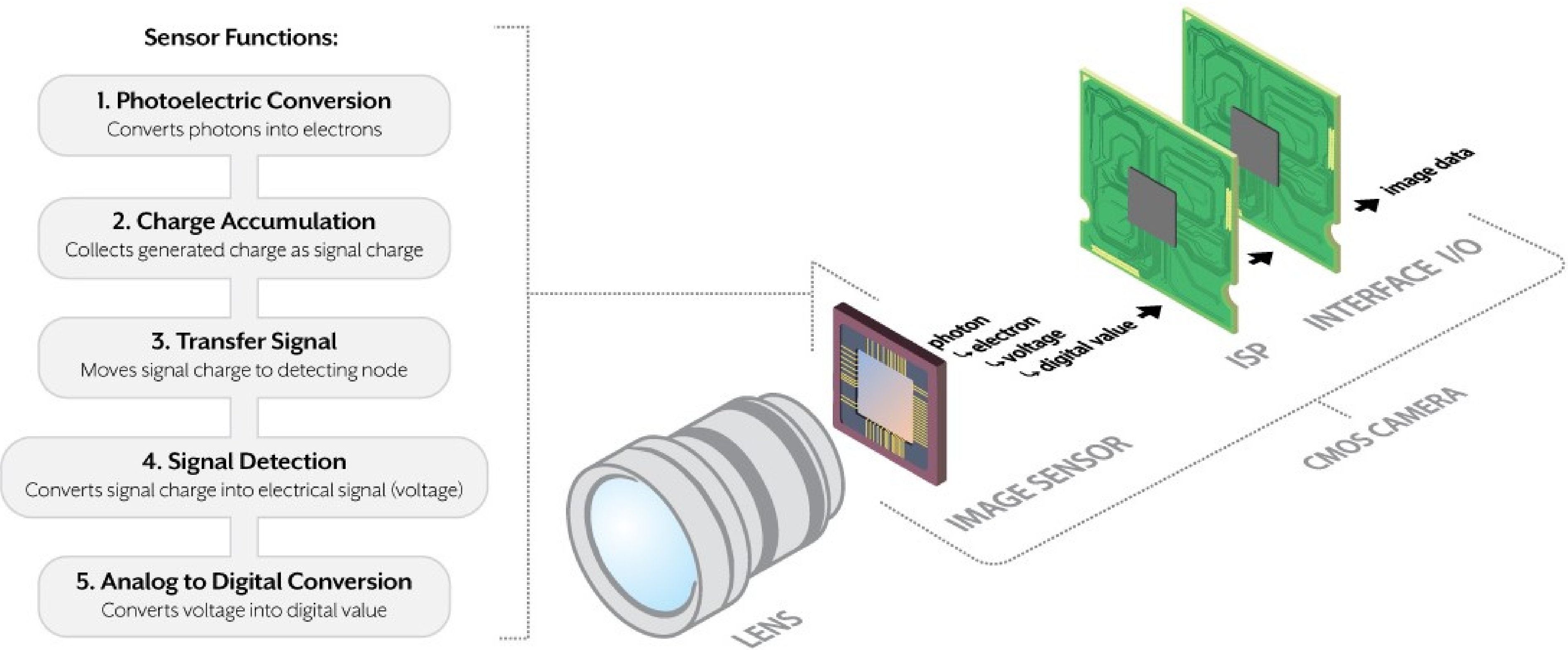






Sensor Functions Inside a Camera

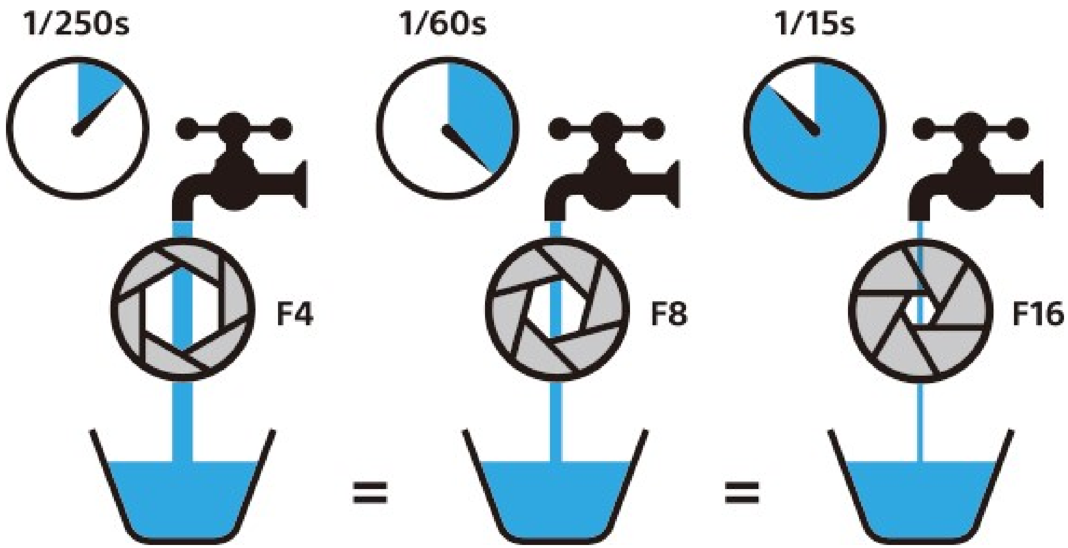
- Typical CMOS Camera Layout



Exposure and Readout

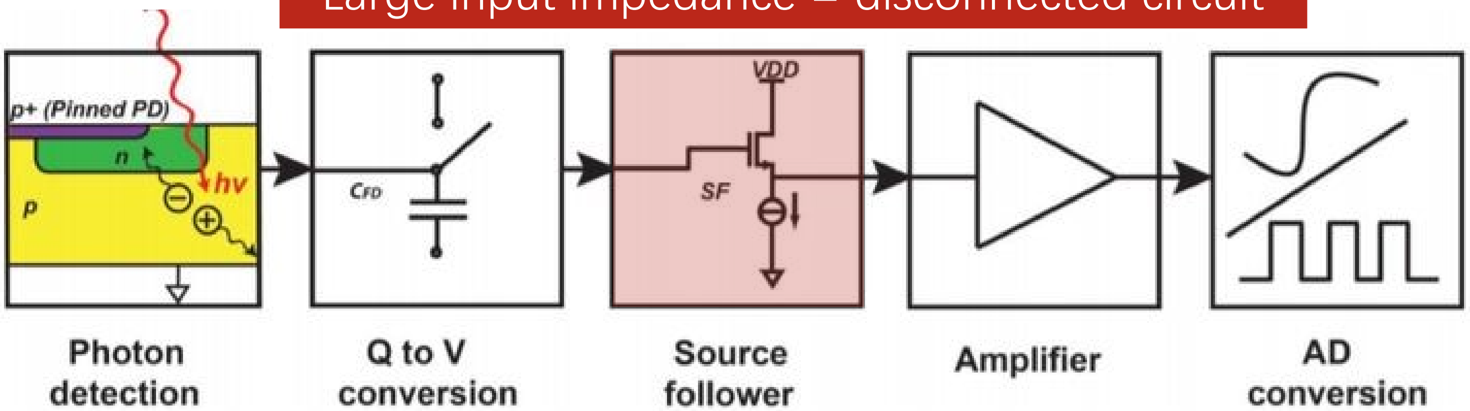
- Analogy for exposure and aperture

How long and how large you turn on the water tap?



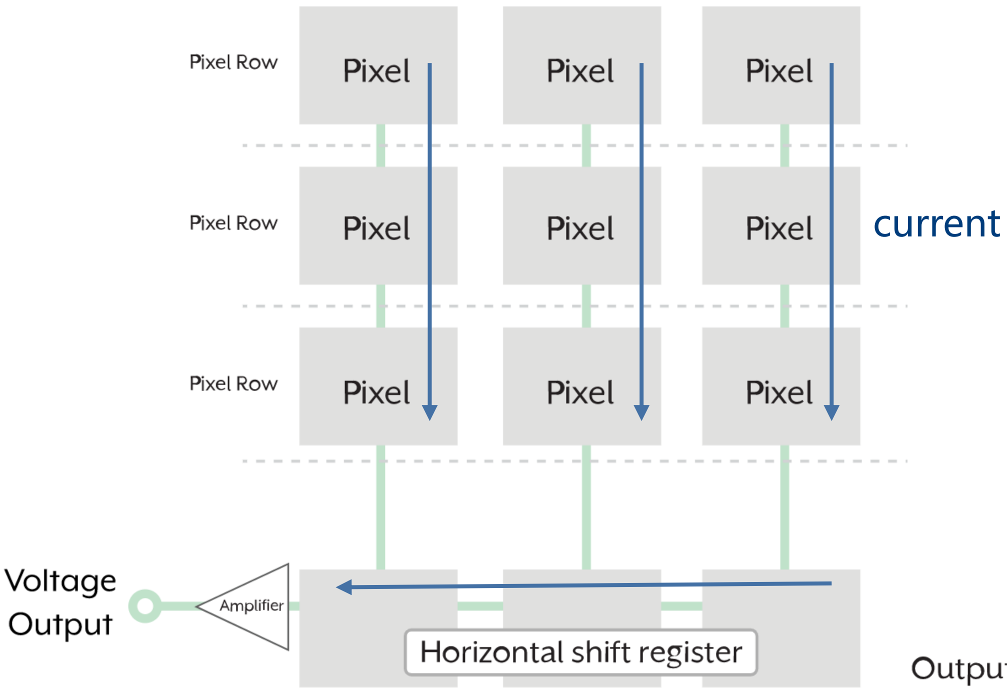
- Readout process

Large input impedance = disconnected circuit

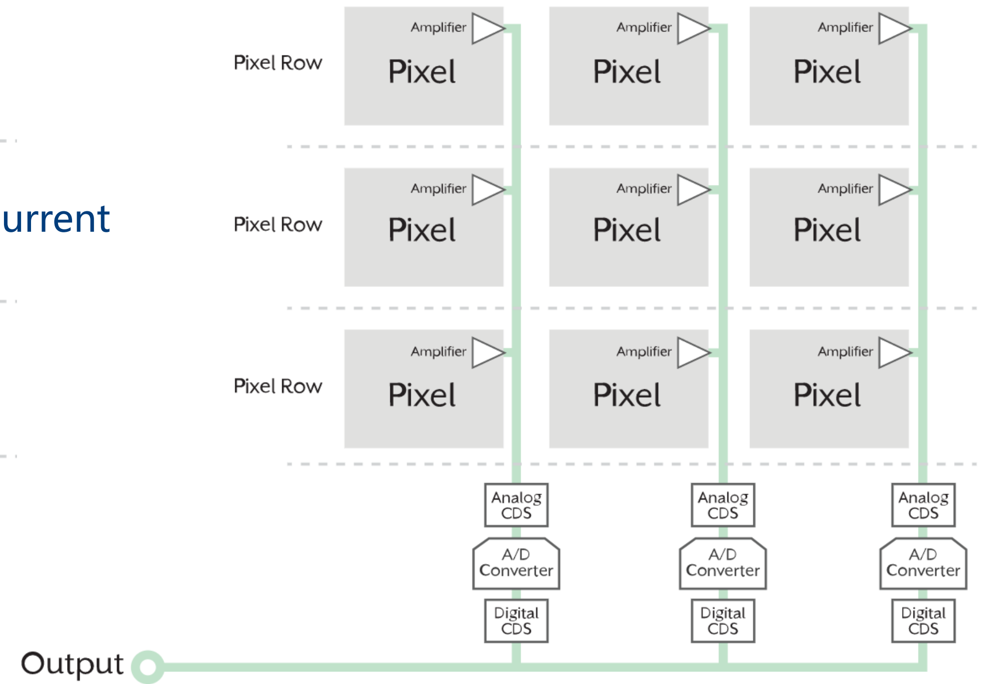


CCD and CMOS Sensors

- The main difference between CCD and CMOS is how they transfer the charge out of the pixel and into the camera's electronics "Read Out"

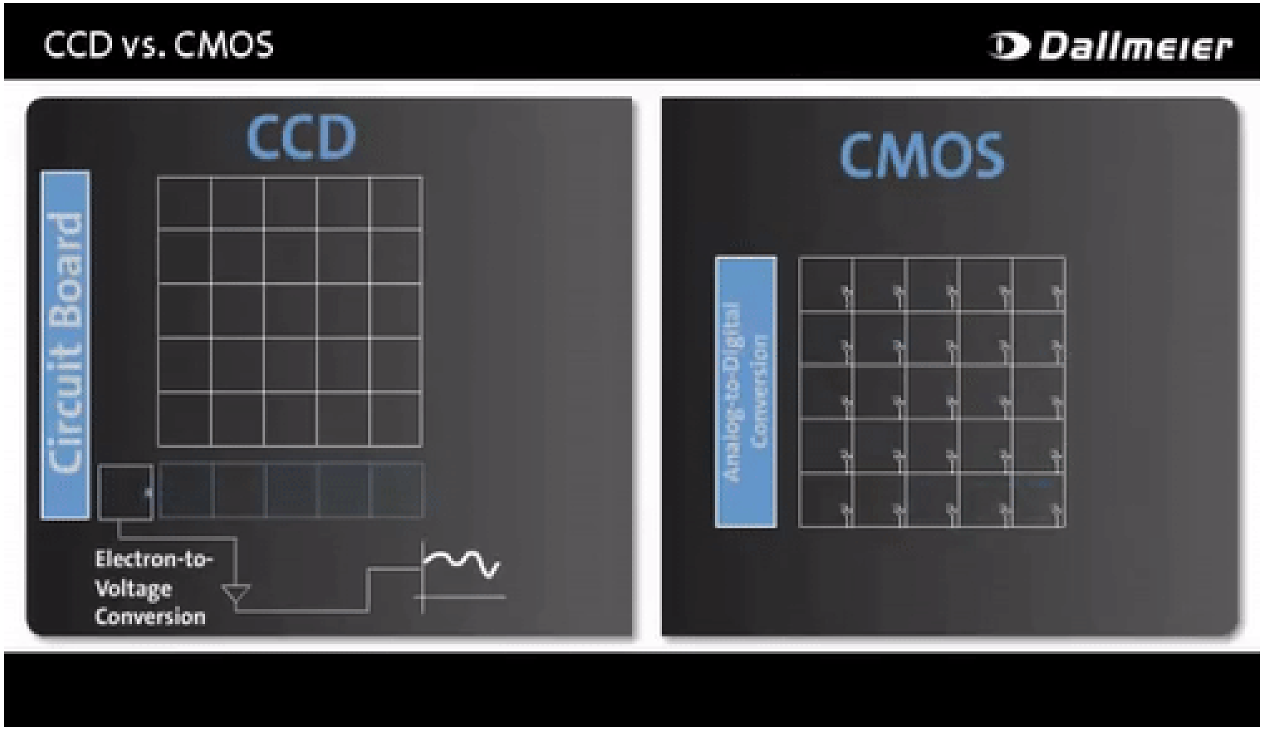


CCD sensors (Charged Couple Device)



CMOS sensors
(Complementary Metal-Oxide Semiconductor)

CCD and CMOS Sensors

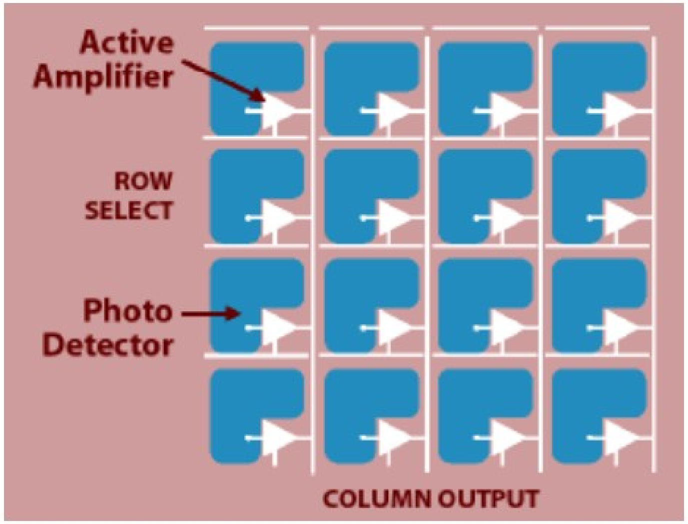


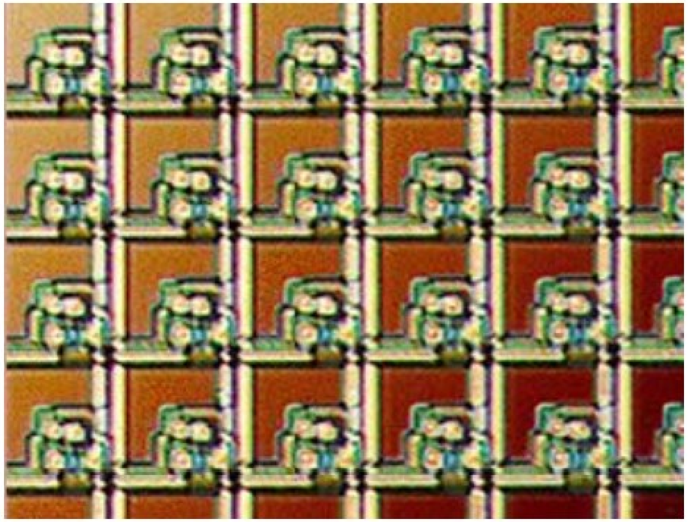
- Global shutter
- Low noise
- Medium range frame rates: charge transfer takes time
- Outputs analog pulse

- Global shutter and rolling shutter
- Low to very low noise
- Very high frame rates
- Outputs completely digital data

CMOS Issues

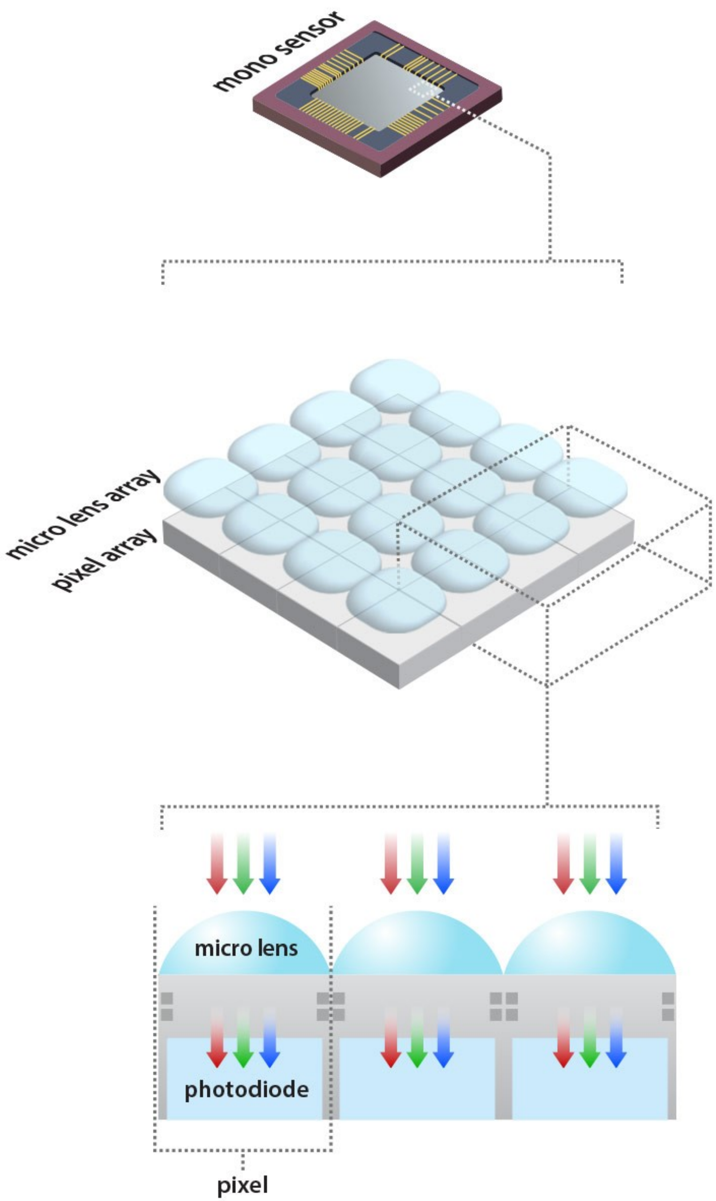
- Pattern Noise
  - Higher image non-uniformities (aka "**Fixed Pattern Noise**") due to unevenness between the individual pixel cells and multiple A/D circuits in column readout.
- Sensitivity
  - CMOS requires areas for amplifiers and other circuits on each pixel, usually resulting in lower sensitivity.

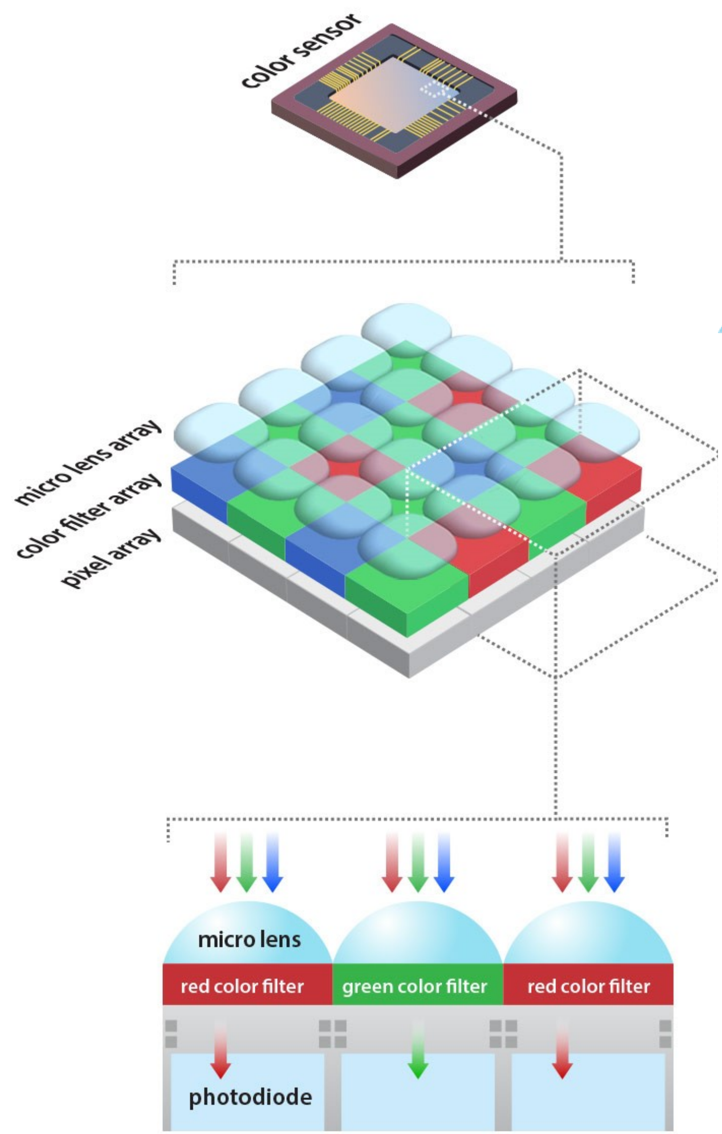


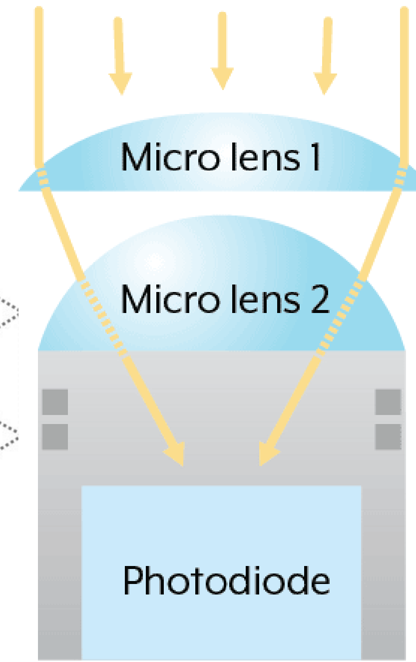


cvp_the-fundamentals-of-camera-and-image-sensor-technology_jon-chouinard.pdf (visiononline.org)

Mono and Color Sensors



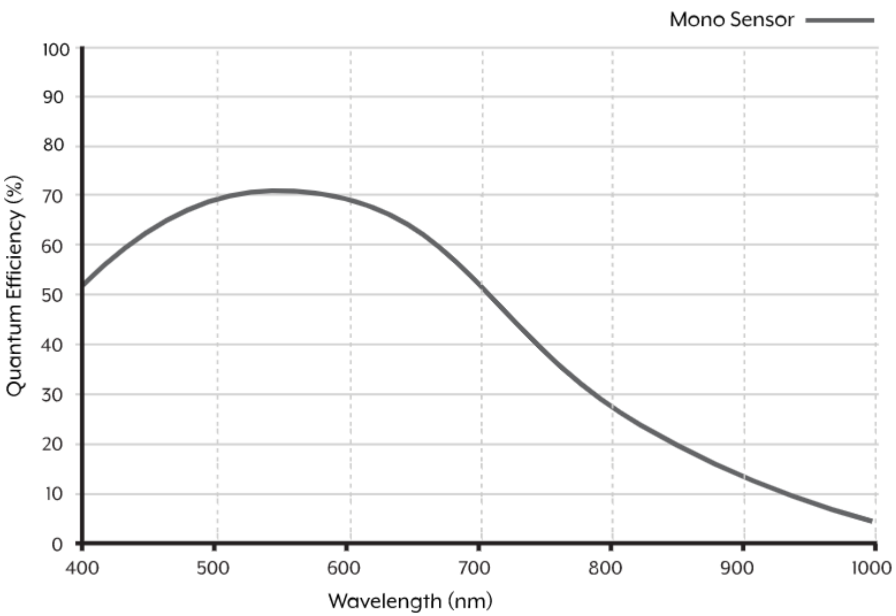


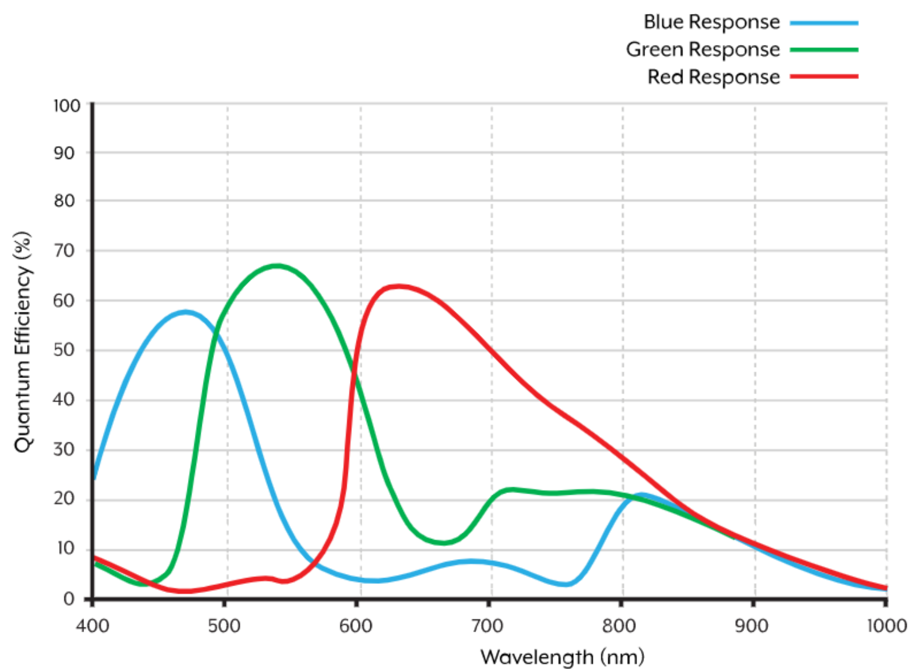


Additional micro lenses are used to help guide photons into the photodiode.

Mono and Color Spectral Response

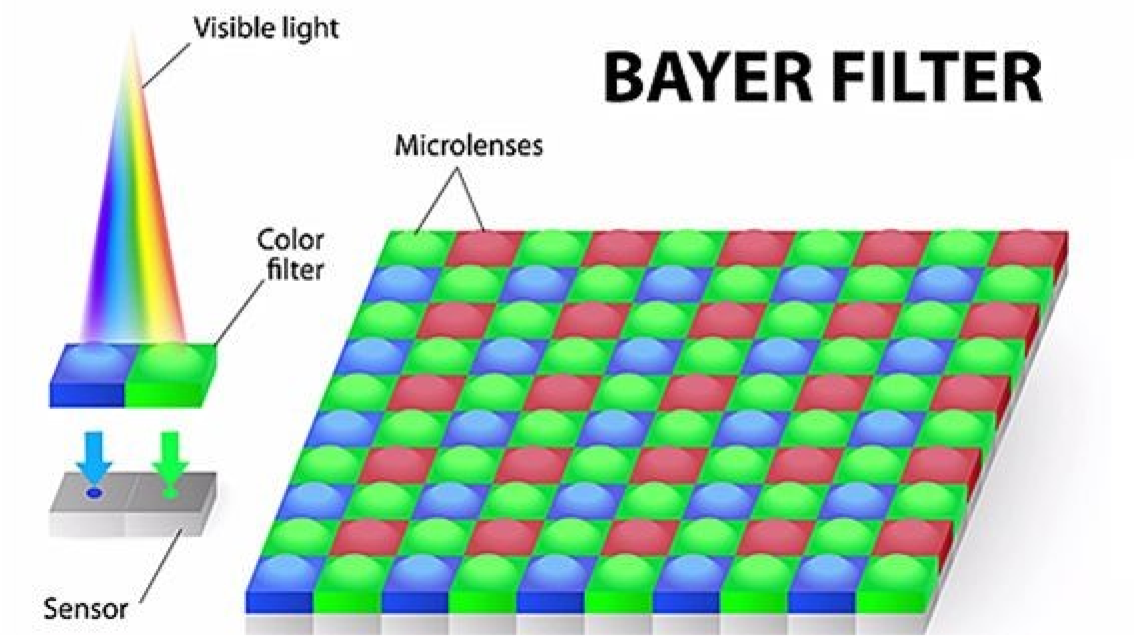
- Color spectral response related to: quantum efficiency of the mono pixel and color filters.

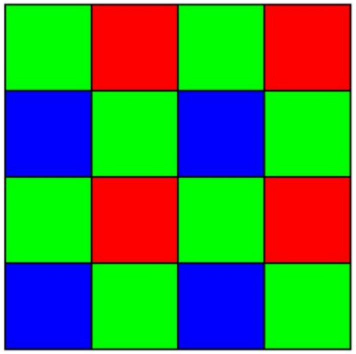


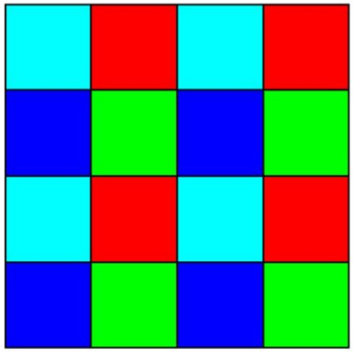


Spectral Response Curves Examples Using Same Sensor Family.
Mono Sensor (Left) and Color Sensor With No IR Cut Filter (Right)

Bayer Color Filter

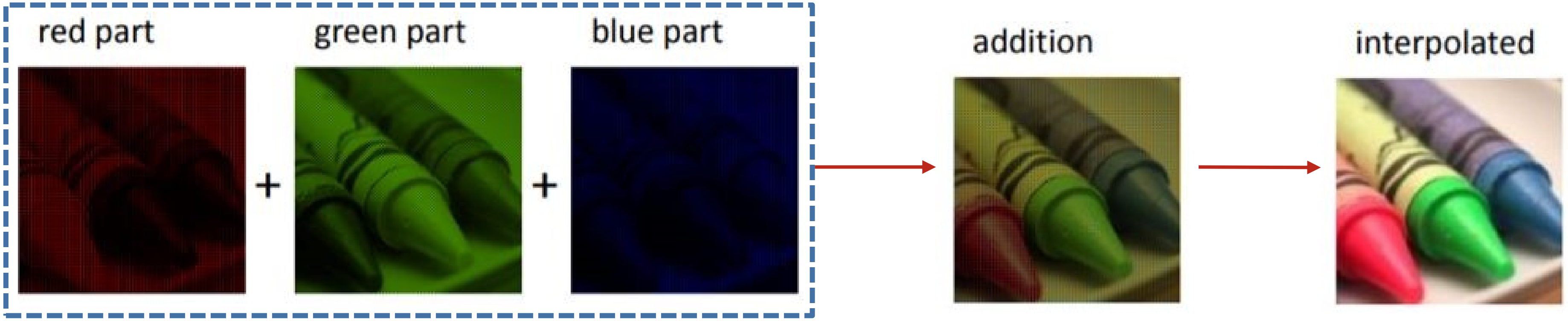






E for "Emerald"

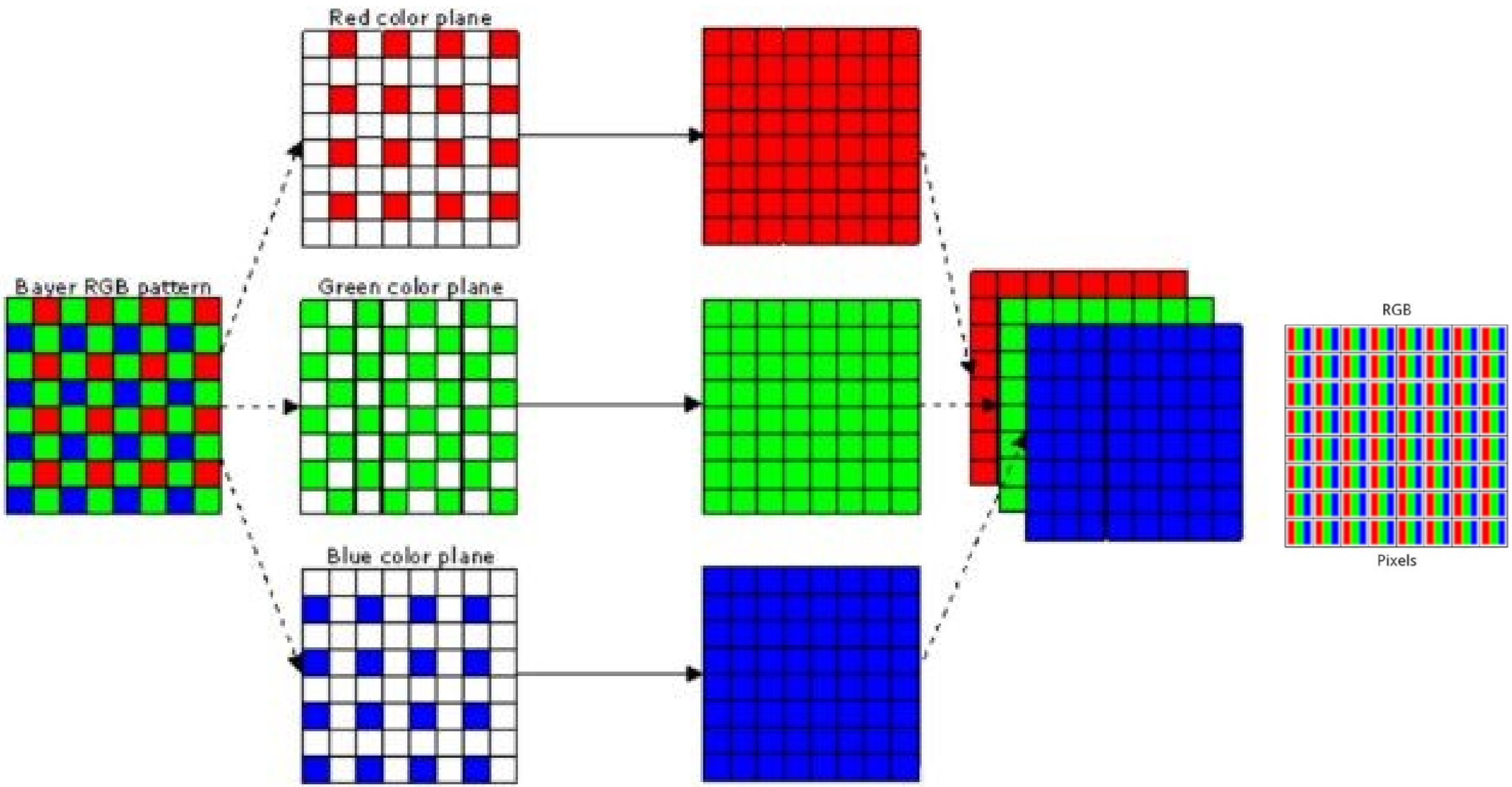
RAW format image



RAW to RGB Interpolation

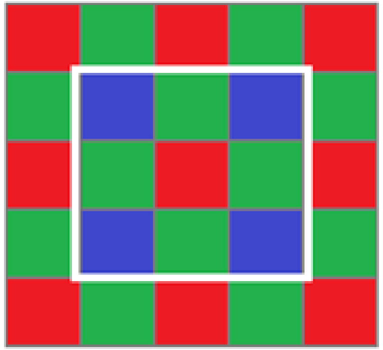
- **ILL-POSED problem**: infinite number of solutions!

"It is art instead of science."

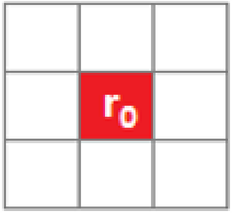


RAW to RGB Interpolation

- One possible interpolation formula

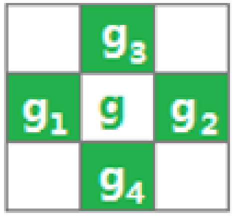


A 3 by 3 bayer window



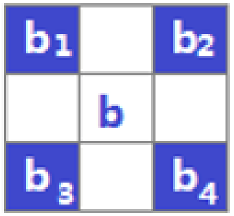
$$r = r_0$$

Interpolating R



$$g = \frac{1}{4}[g_1 + g_2 + g_3 + g_4]$$

Interpolating G

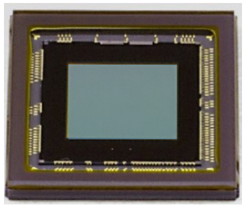


$$b = \frac{1}{4}[b_1 + b_2 + b_3 + b_4]$$

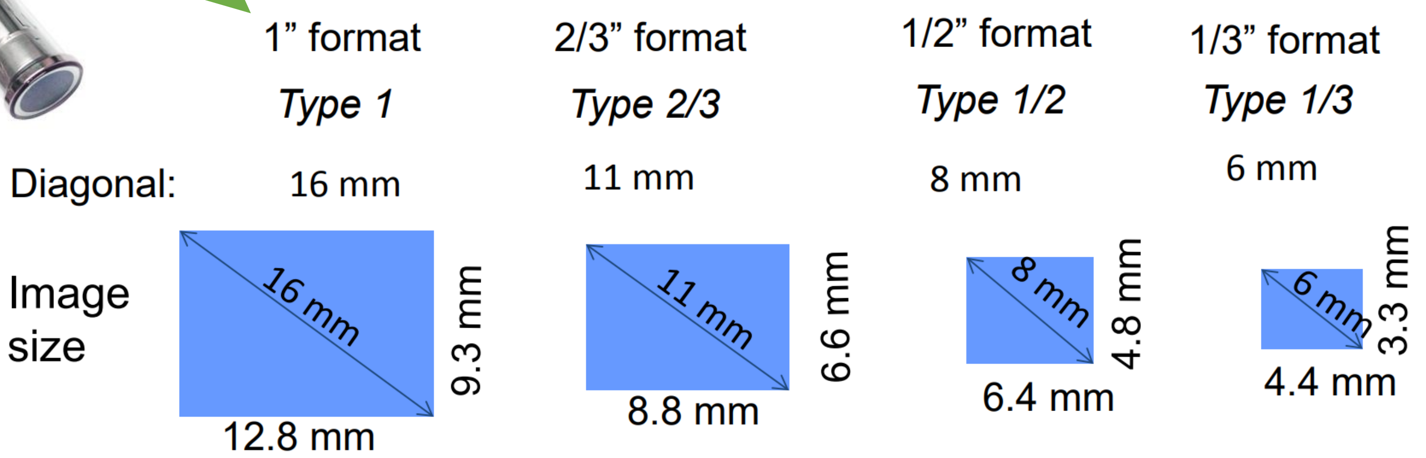
Interpolating B

Image Sensor Format (Size)

- Image sensors come in different *format types* (also known as *optical class*, *sensor size* or *type*) and packages.
- Knowing the sensor format is **essential for lens selection**.
- Note that: sensor formats only describe the area of the **sensor chip** and not the entire sensor package.



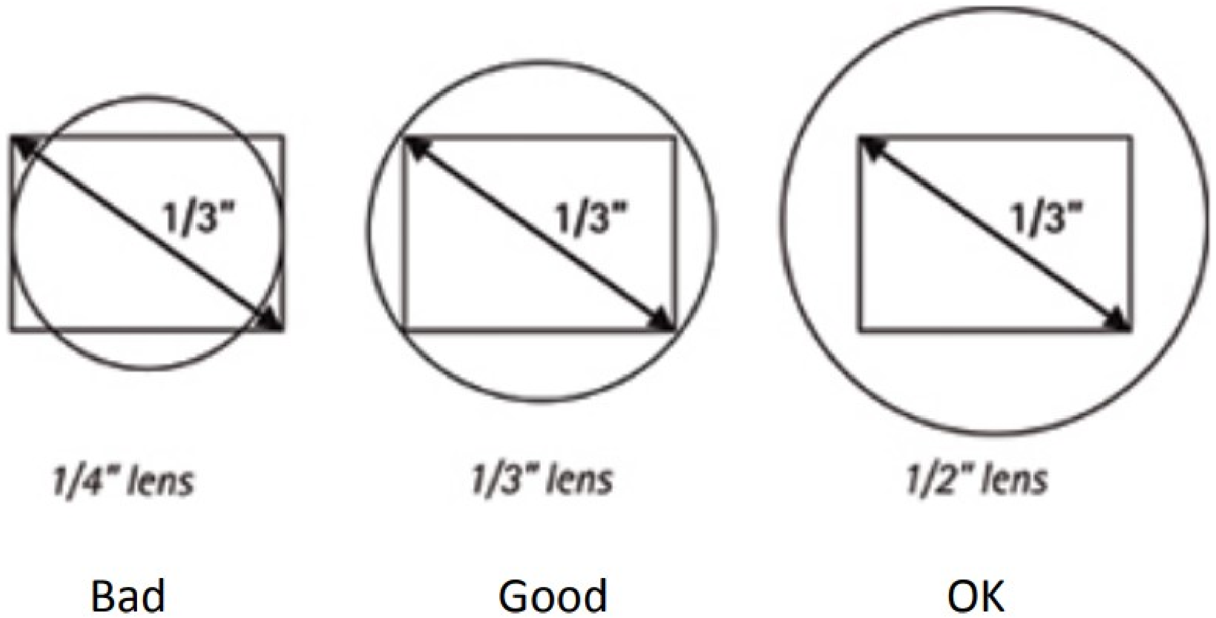
The number indicates the diameter of old video camera tubes!



Most common image formats.

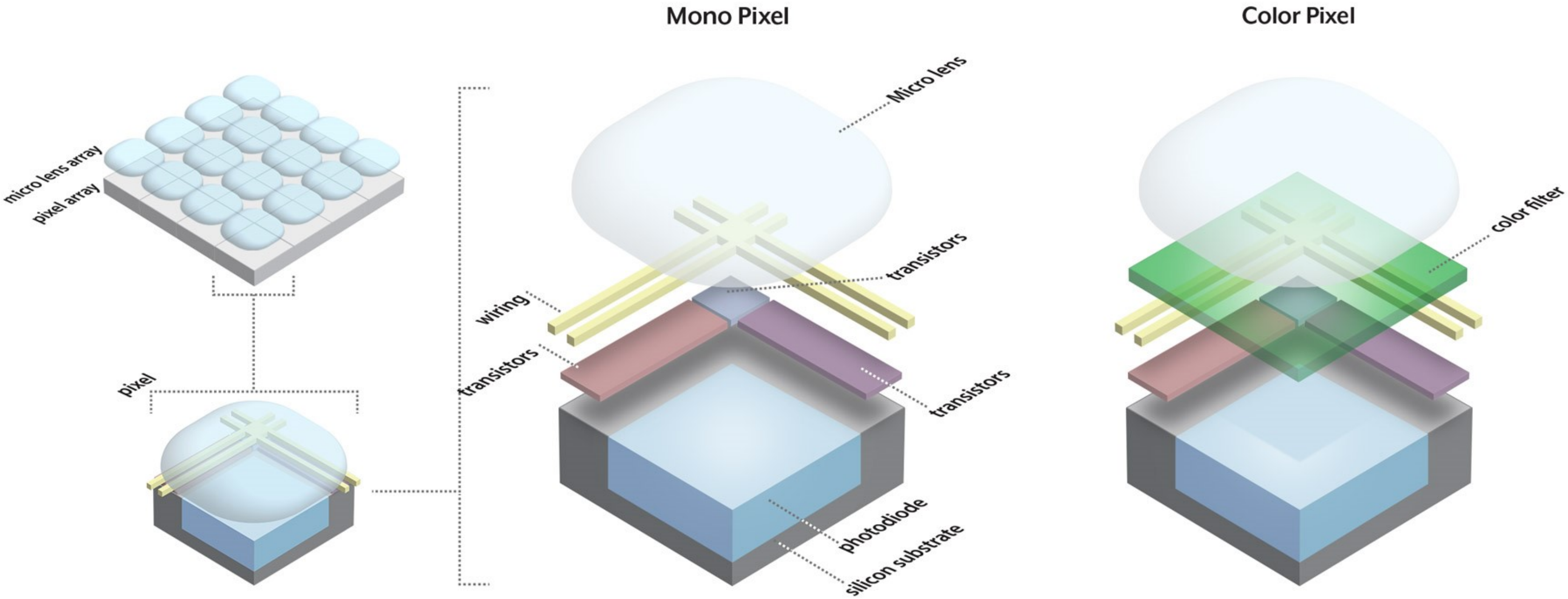
Image Sensor Format (Size)

- Why is sensor format essential for lens selection?
- Lens specs must match image format
  - 1/3″ Sensor Format ideally should be paired with 1/3″ format lens
  - Larger format lens can be used on smaller sensor
  - Smaller format lens can NOT be used on larger sensor



Sensor Pixel Size

- Pixel size is measured in micro-meters (µm) and includes the entire area of both the **photodiode** and **surrounding electronics**.

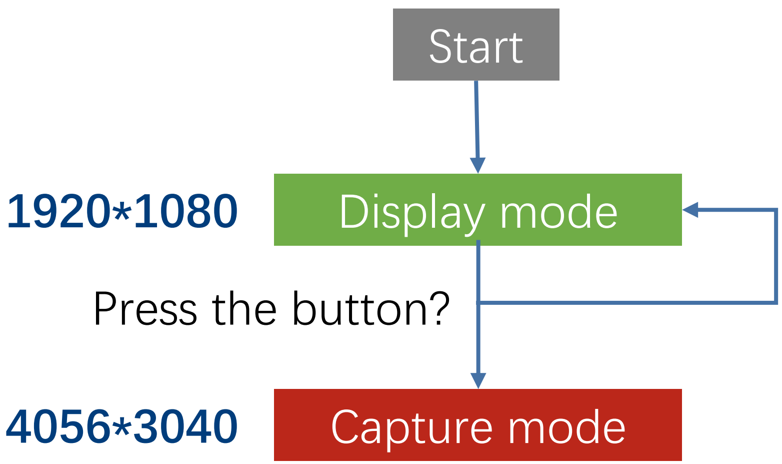


A simplified layout of a CMOS mono and color pixel.

- Larger pixel size better for **increased light sensitivity**, but also results in **larger sensor size.**

Resolution

- Maybe the most important and straightforward parameter!
- Different resolutions in capture mode and display mode:

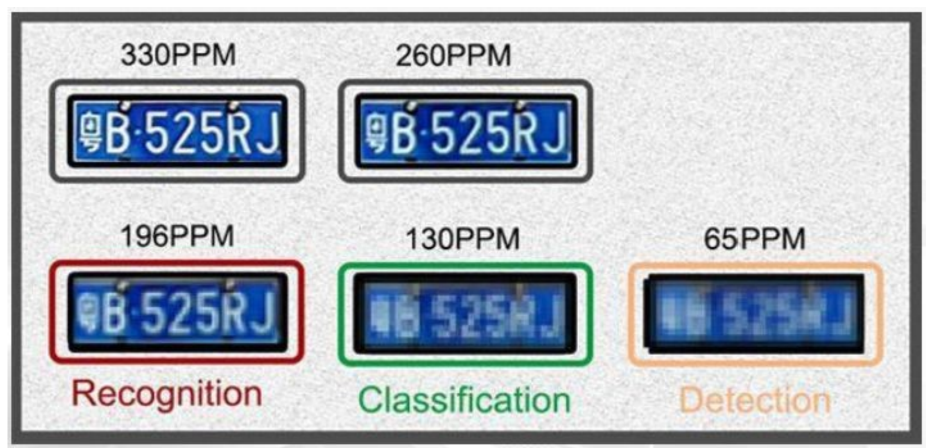


- Why difference resolutions in different mode?
  Do we really need high resolution in display mode?

Resolution

- Five levels of image quality according to pixel density:
  - **Identification → Recognition → Classification → Detection → Useless**

The **PPM** (Pixels Per Meter) value defines the **number of points** or divisions for **each meter of the subject** that the camera is imaging.



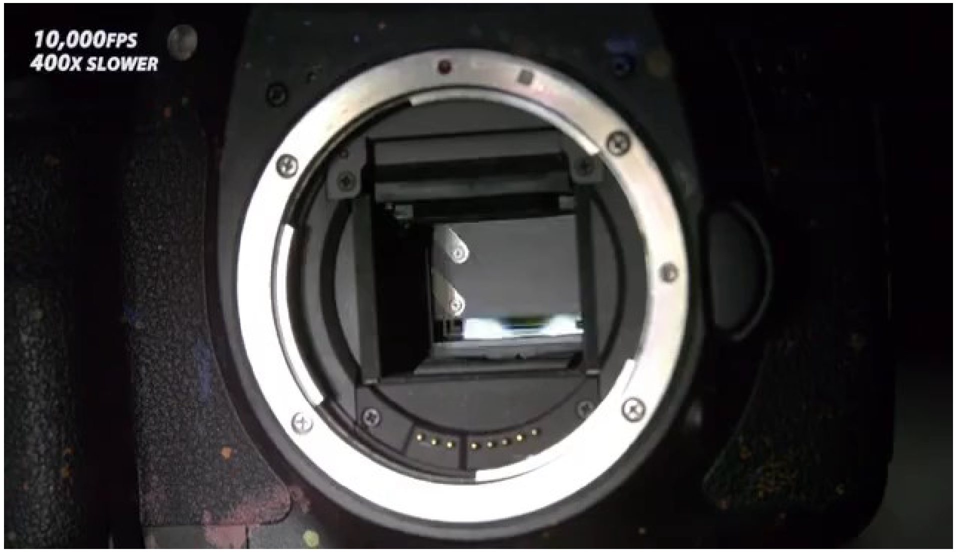
Resolution: Exercise

- The width of face is 200mm. The pixel density = 400 PPM.
  Find out the **face width** in pixel numbers.

- PPM is varying w.r.t. camera-object distance.

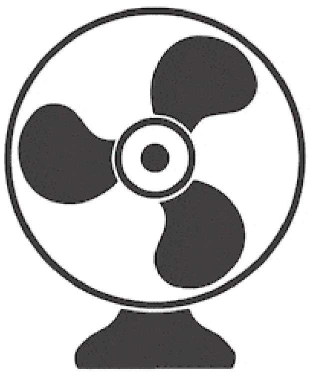
Global versus Rolling Shutter

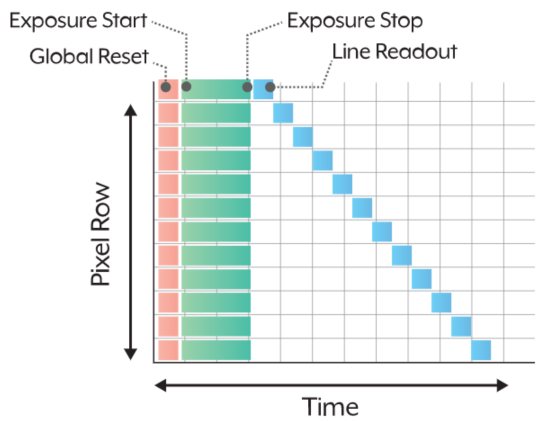
- From mechanical shutter to electronic shutter

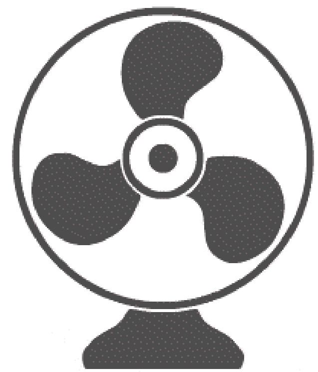


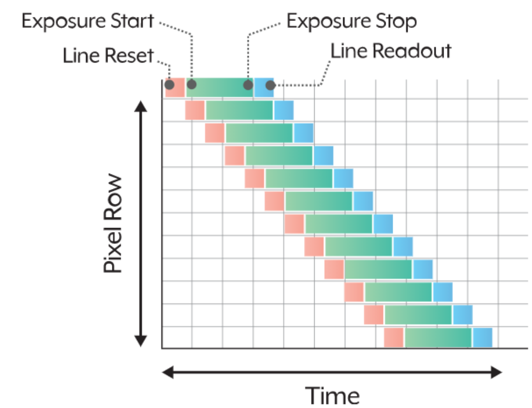
- A **global shutter** allow exposure of the whole frame at the same time.
- A **rolling shutter** will expose the frame line after line. The number of exposures equals the number of lines in the frame.

Global versus Rolling Shutter









- **No image distortion**
- **Requires large readout bandwidth**

- **Distortion for fast-moving objects**
- **Lower readout bandwidth**
- **Faster framerate**

Global versus Rolling Shutter

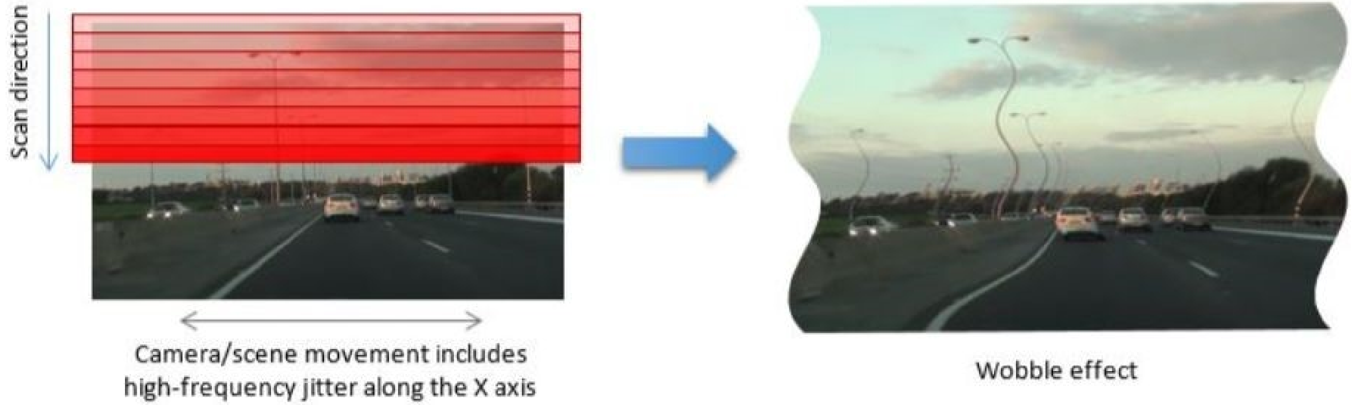
- Both global and rolling shutters have the **motion blur** issue

  - **Sometimes, we need the blur!**

Global versus Rolling Shutter

- Rolling shutter effects
  - Skew

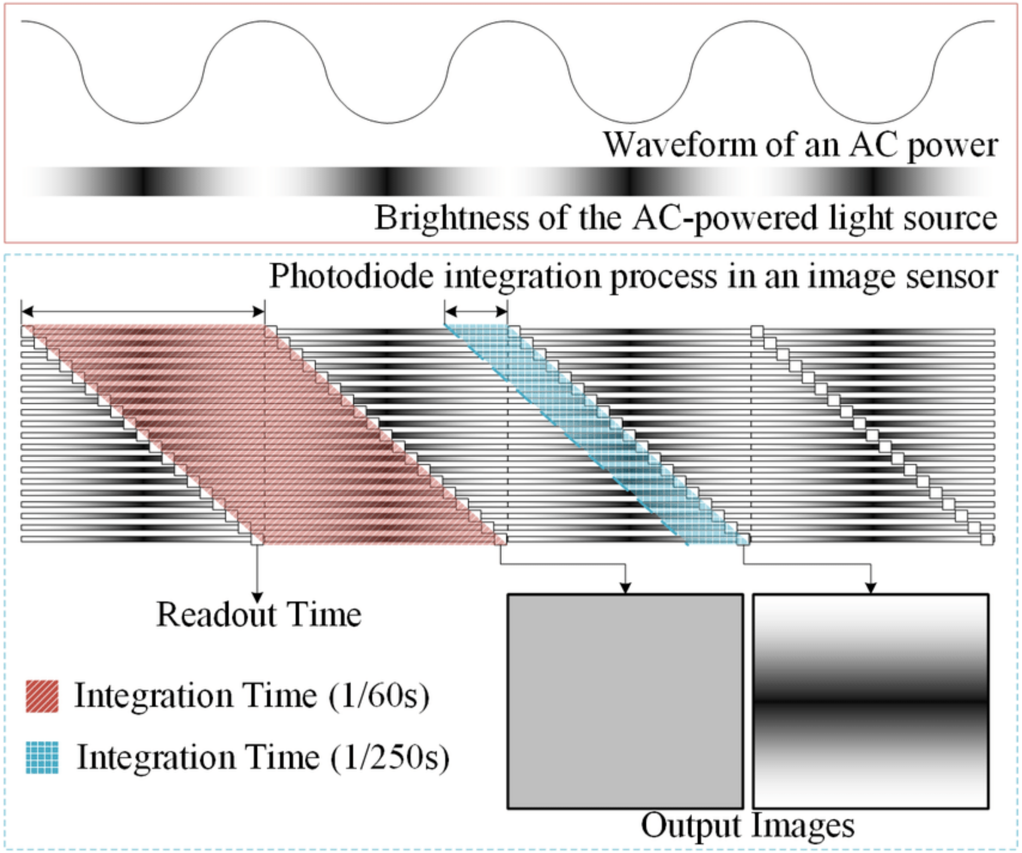
  - Wobble



  - Partial flash

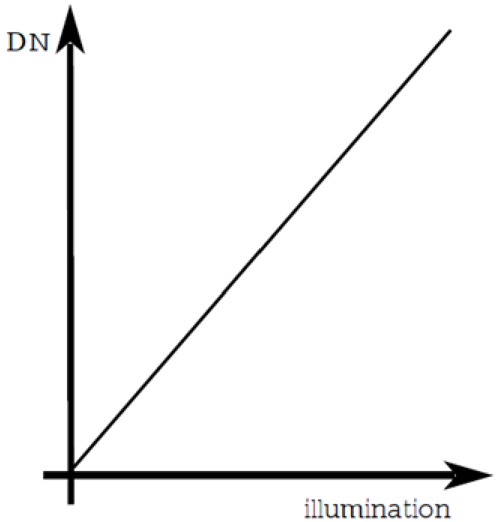
Flickering

- Usually happens if the exposure time is not an integer multiple of light source period.

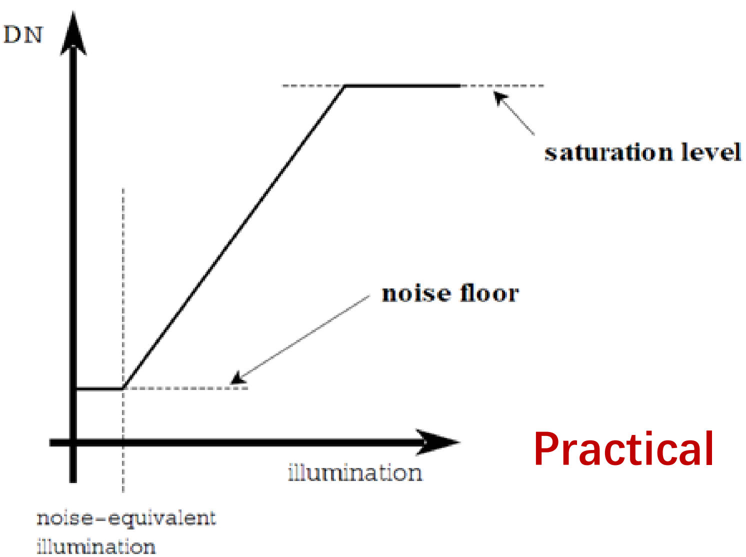


Input-output Characteristics

- Ideal input-output relation:
  - Linearity
  - Infinite input/output range
  - High sensitivity: small input variation leads to large output variation
  - High framerate
  - High resolution
  - Low power consumption
  - Simple manufacturing process
  - Low cost



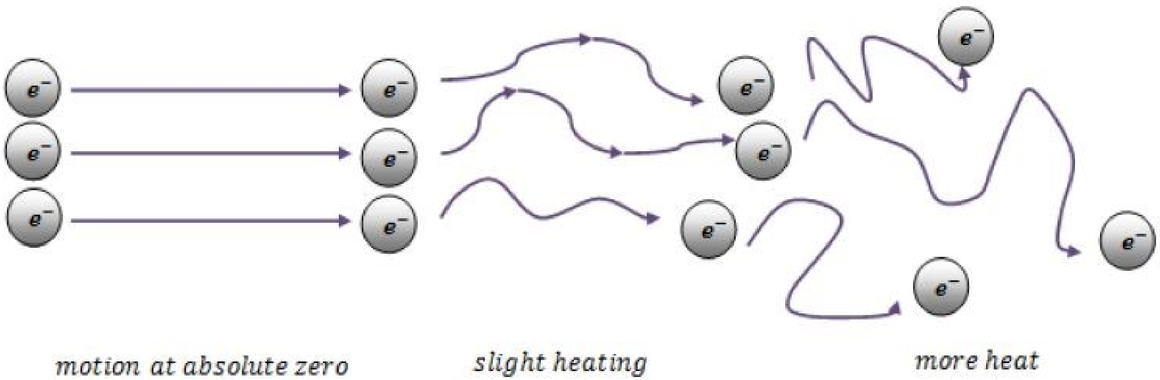
Ideal

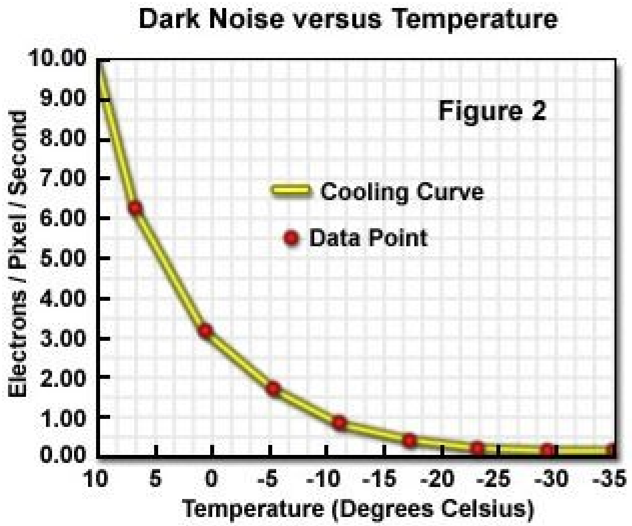


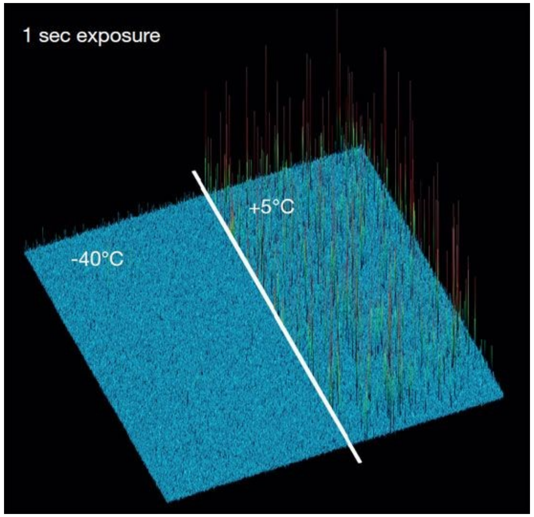
Practical

Noises

- Dark noise
  - Random dark current due to **electron's thermal motion**



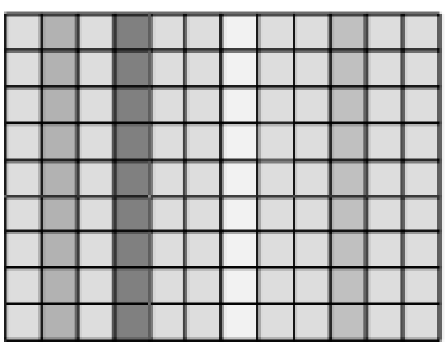




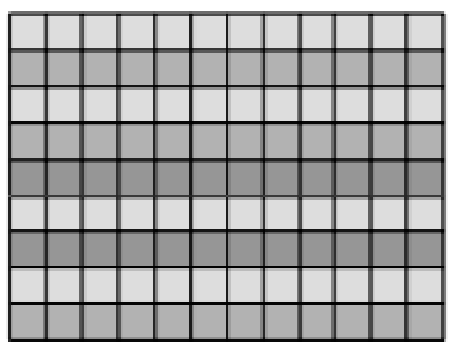
Noises

- Quantization noise
  - Brought in during ADC, due to finite accuracy of digital signal
  - Digital signals are usually 10-14 bits after ADC

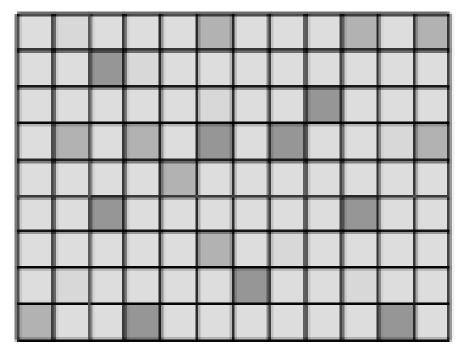
- Fixed Pattern Noise (FPN)
  - Due to the spatial nonuniformity of the pixels
  - Can be corrected in calibration



Column FPN



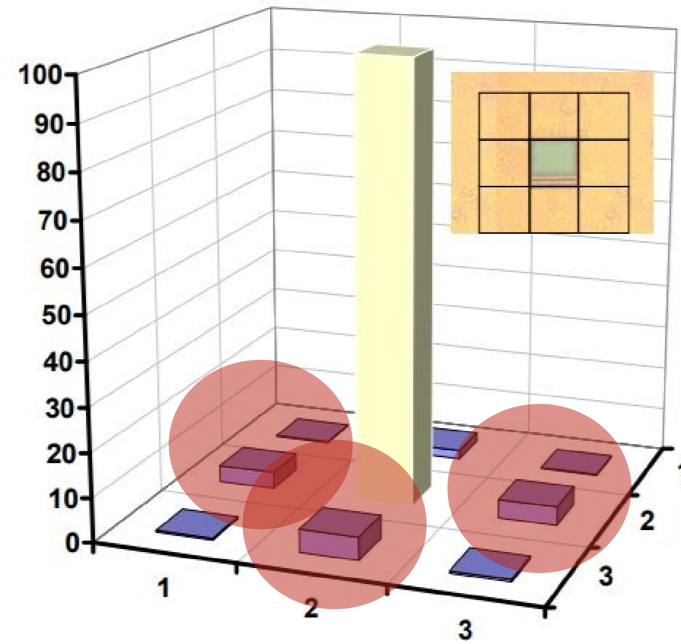
Row FPN



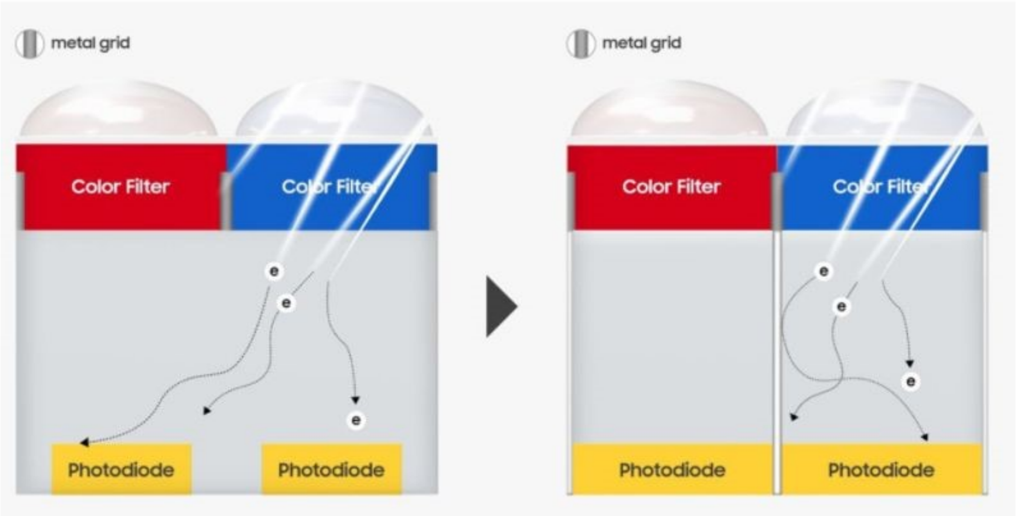
Pixel FPN

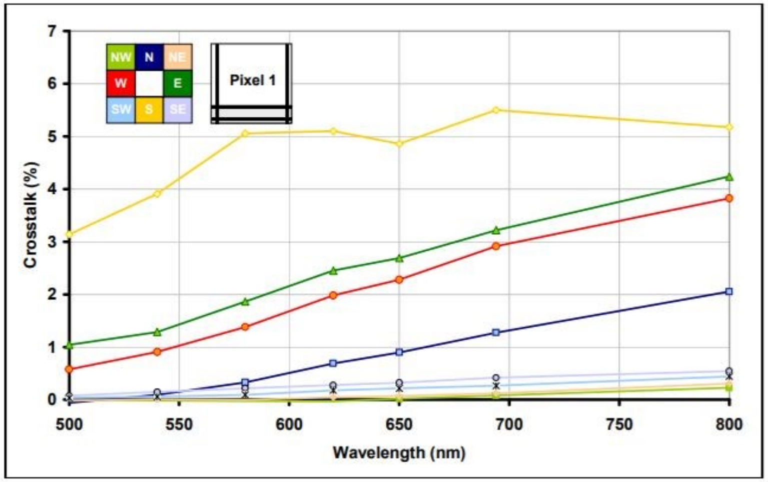
Noises

- Crosstalk
  - Photon captured by adjacent pixels
  - Red areas should have “zero output” →
  - Related to wavelength
  - Reduce the crosstalk:
    e.g., ISOCELL by Samsung



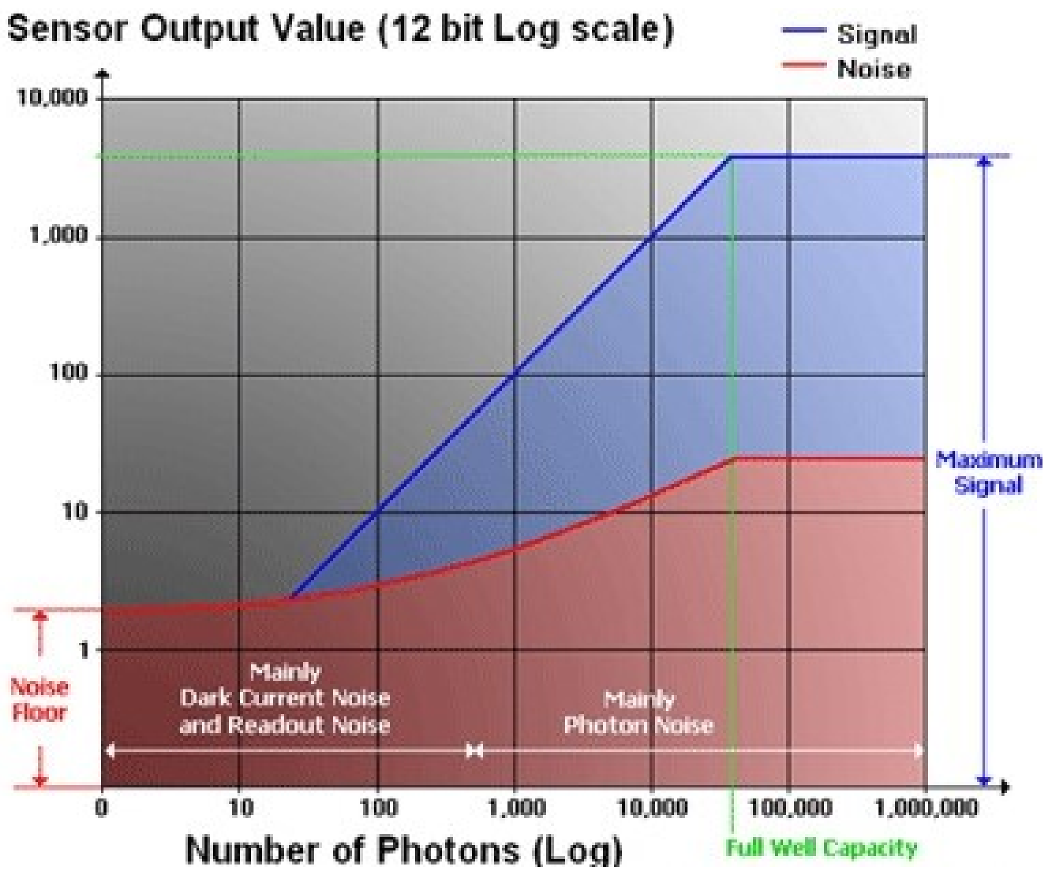
The metal grid prevents electrons from entering the adjacent photodiodes





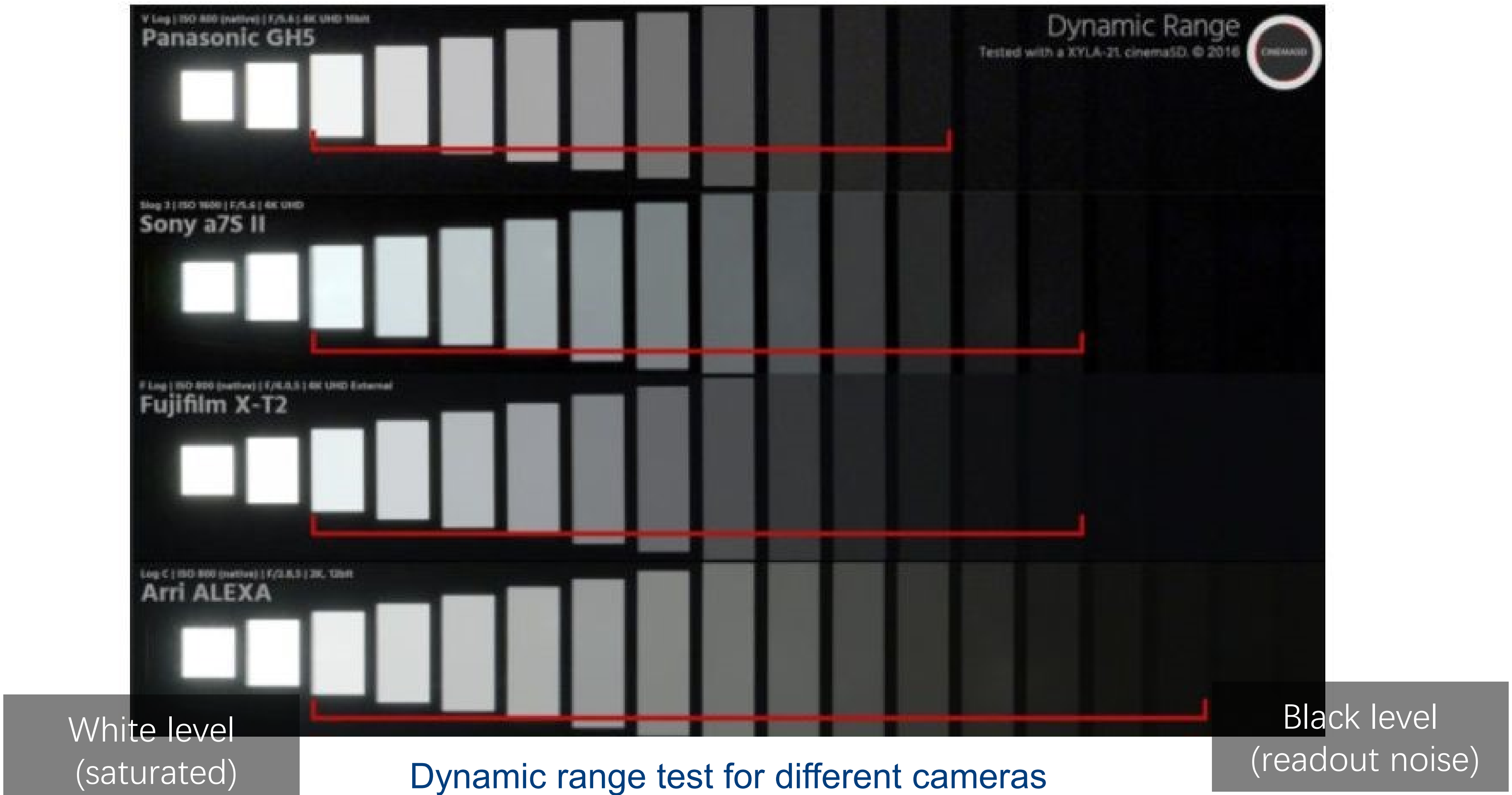
Signal-Noise Ratio (SNR)

- Main criterion for image fidelity evaluation
- Defined as **the ratio of the input signal power to the input referred noise power**, and is a function of input signal. It is typically measured in dBs.



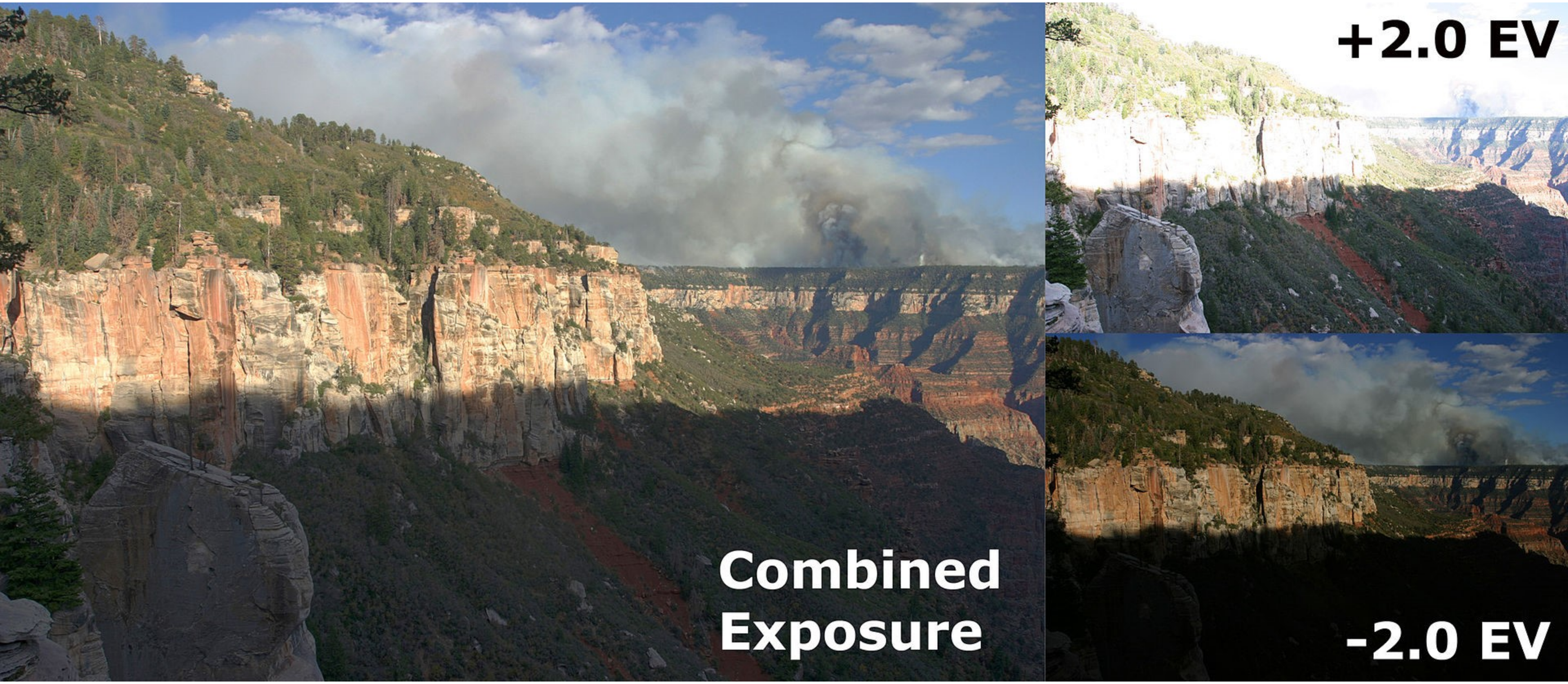
Dynamic Range

- The ratio **largest possible signal** to the **signal that is just at the noise level** in one image.



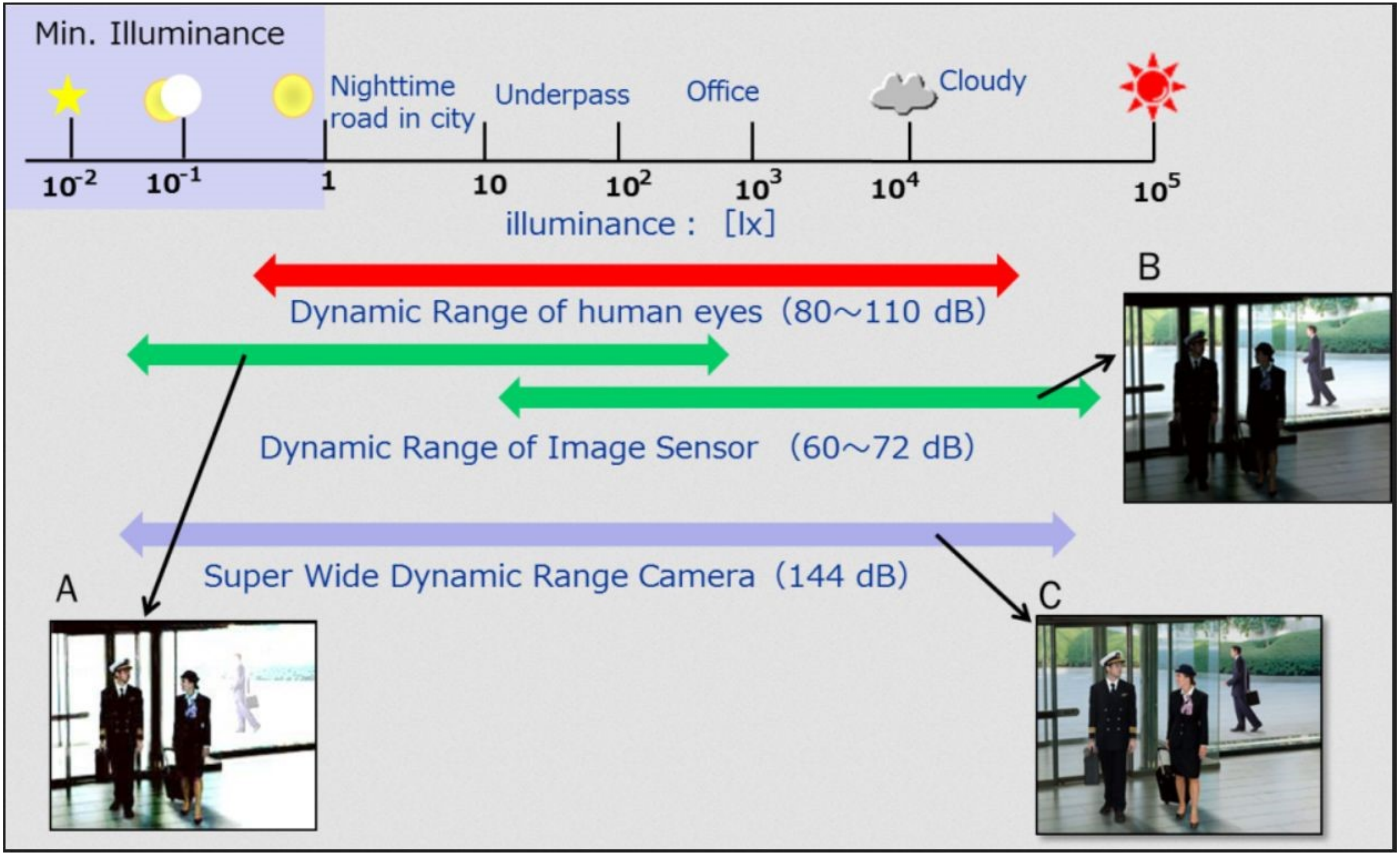
Dynamic range test for different cameras

Dynamic Range



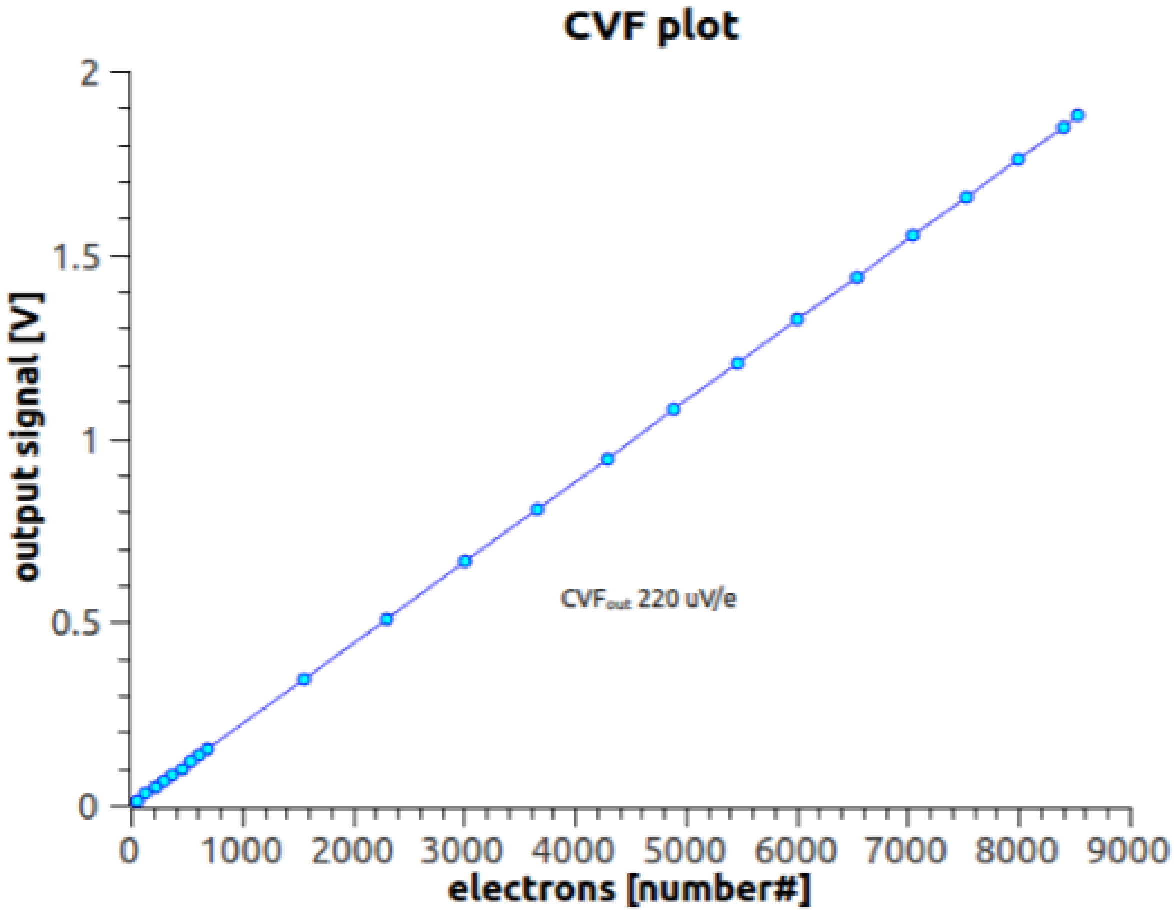
Dynamic Range

- Dynamic ranges for typical scenarios



Sensitivity

- Photons to Electrons: **Quantum Efficiency**
- Electrons to Output signal: **Charge-Voltage Factor (CVF)**



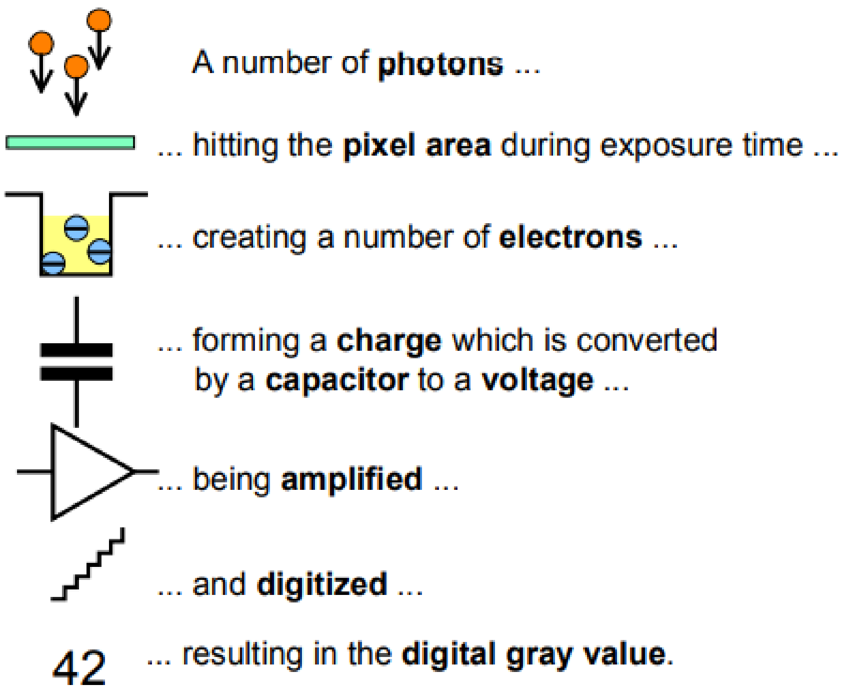
Sensitivity: Demo

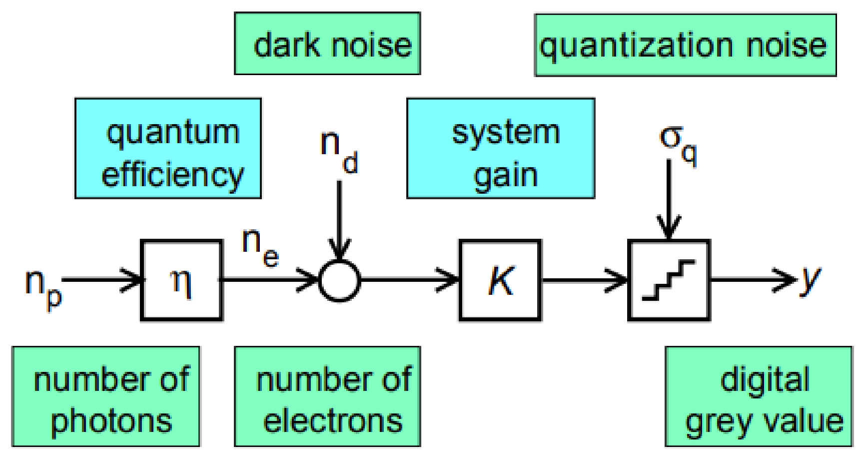
- How does the sensitivity influence the output images?

BROOKMAN TECHNOLOGY

EMVA Standard 1288

- How to (mathematically) model the process?
  - Standard for Characterization of Image Sensors and Cameras
  - Issued by European Machine Vision Association

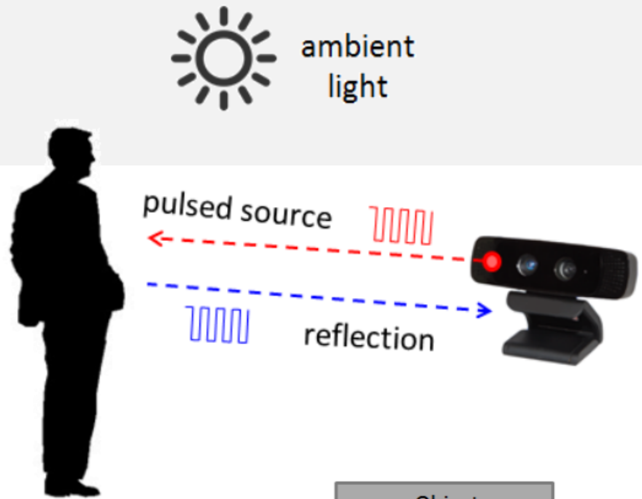




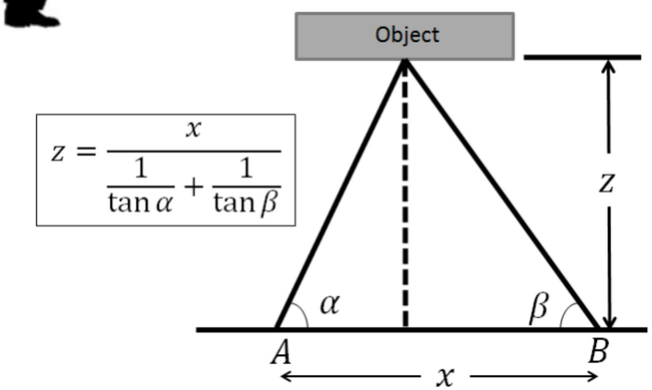
Physical model of the camera (left), and Mathematical model of a single pixel (right).

3D Imaging Technologies

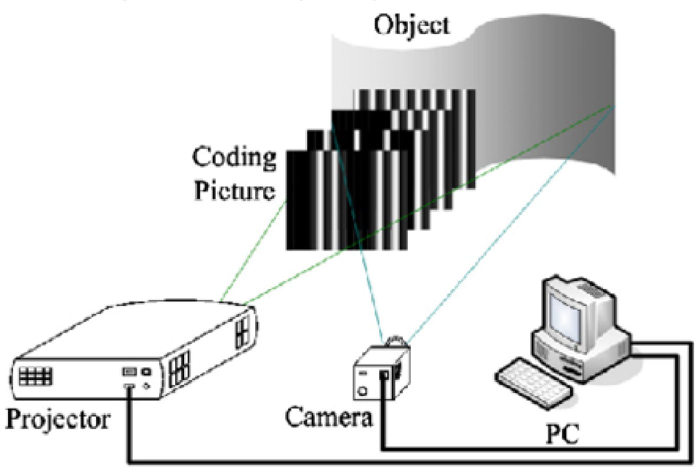
- TOF cameras

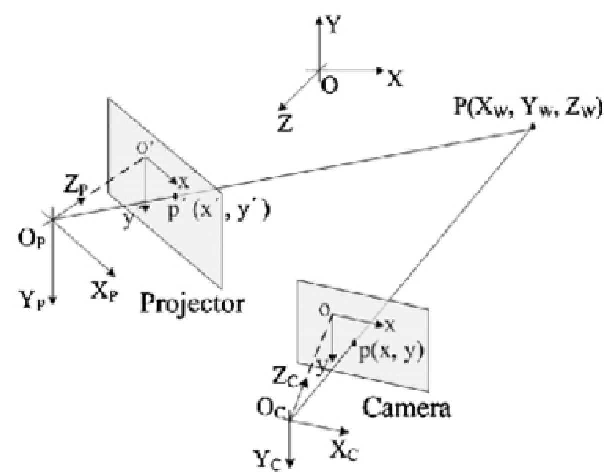


- Stereo vision



- Structured light imaging





3D Imaging Technologies

Comparison of 3D Imaging Technologies

CONSIDERATIONS	STEREO VISION	STRUCTURED-LIGHT	TIME-OF-FLIGHT (TOF)
Software Complexity	High	Medium	Low
Material Cost	Low	High	Medium
Compactness	Low	High	Low
Response Time	Medium	Slow	Fast
Depth Accuracy	Low	High	Medium
Low-Light Performance	Weak	Good	Good
Bright-Light Performance	Good	Weak	Good
Power Consumption	Low	Medium	Scalable
Range	Limited	Scalable	Scalable
APPLICATIONS			
Game		X	X
3D Movies	X		
3D Scanning		X	X
User Interface Control			X
Augmented Reality	X		X

3D Imaging Data Representation

- Depth map
- Point cloud

Event Cameras / Dynamic Vision Sensors

- Challenges in conventional image sensors

The past 60 years of research have been devoted to frame-based cameras ...but they are not good enough!

Latency & Motion blur

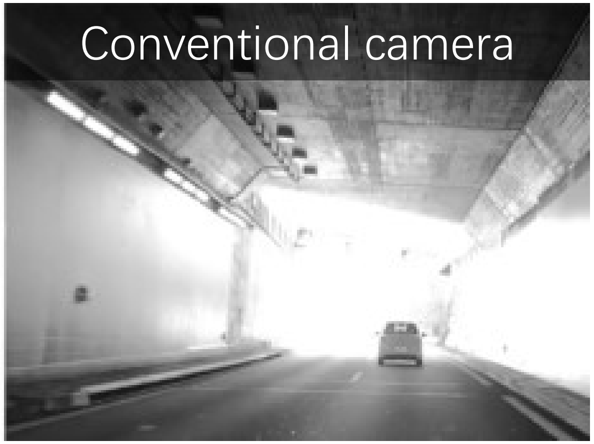
Dynamic Range

Event cameras do not suffer from these problems!

http://rpg.ifi.uzh.ch/docs/scaramuzza/Tutorial_on_Event_Cameras_Scaramuzza.pdf

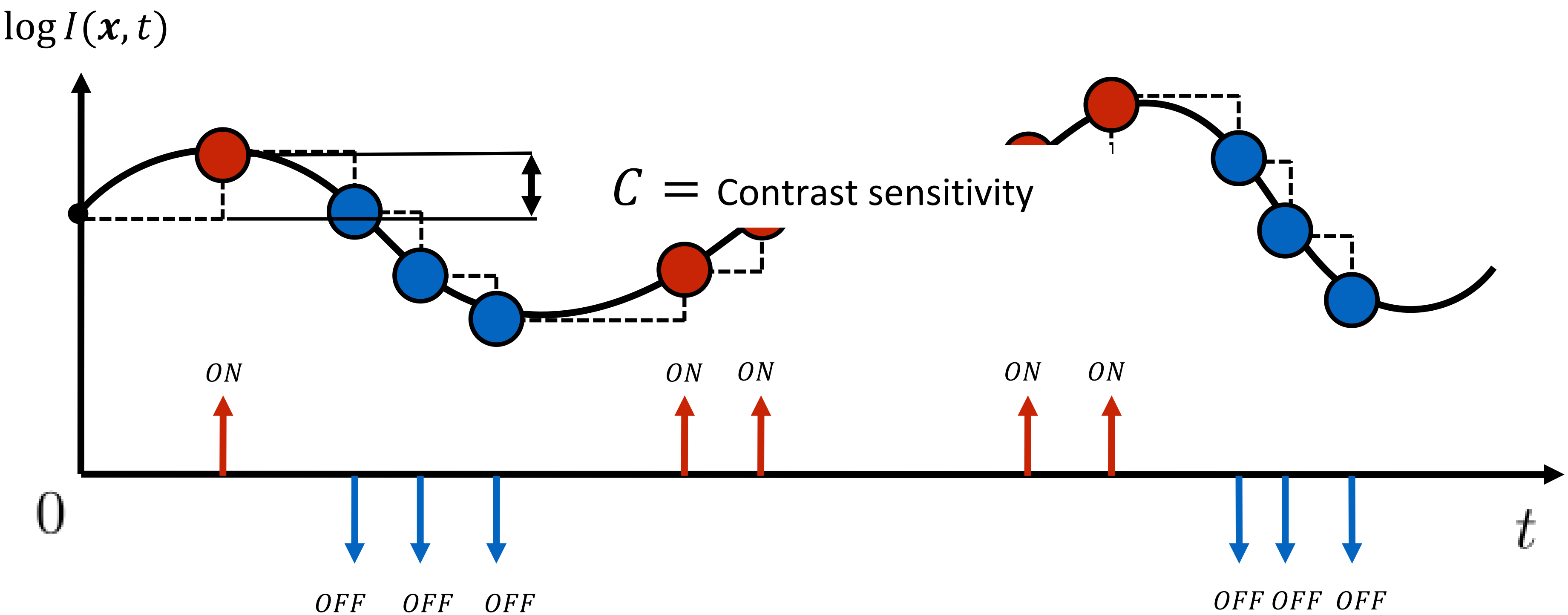
Event Cameras / Dynamic Vision Sensors

- Event-based high speed video reconstruction and HDR scene reconstruction



Event Cameras / Dynamic Vision Sensors

- Event Triggering Principle
  - Each pixel is independently triggered



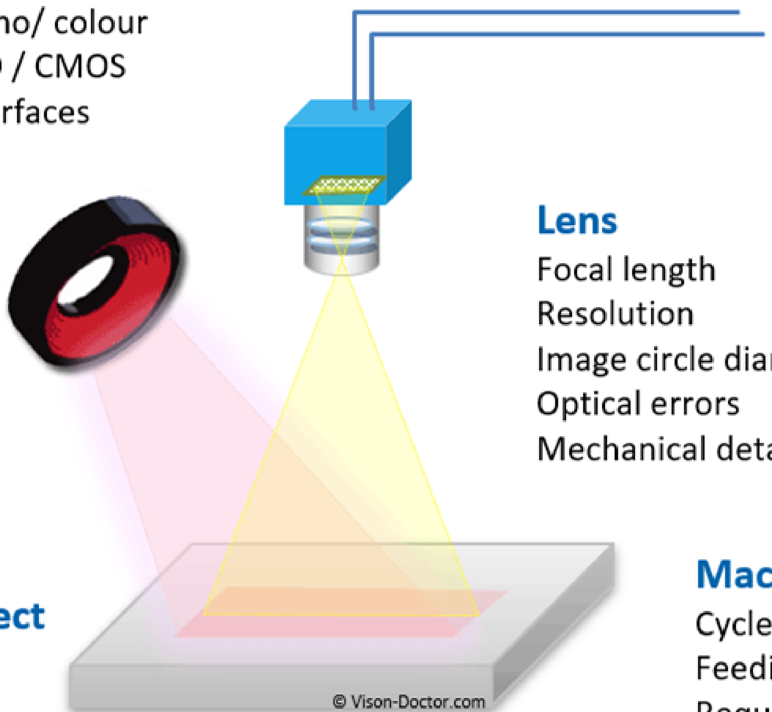
Events are triggered **asynchronously!**

Machine Vision Systems

- Enough for sensor... How to build a machine vision system?

Camera & sensor
Resolution
Sensor size
Mono/ colour
CCD / CMOS
Interfaces

Machine vision system
Software algorithms
Speed
Interfaces
Programming Interface
HMI



Illumination
Form factor
Working distance
Color of light
Direct/ diffuse light
Intensity
Alignment

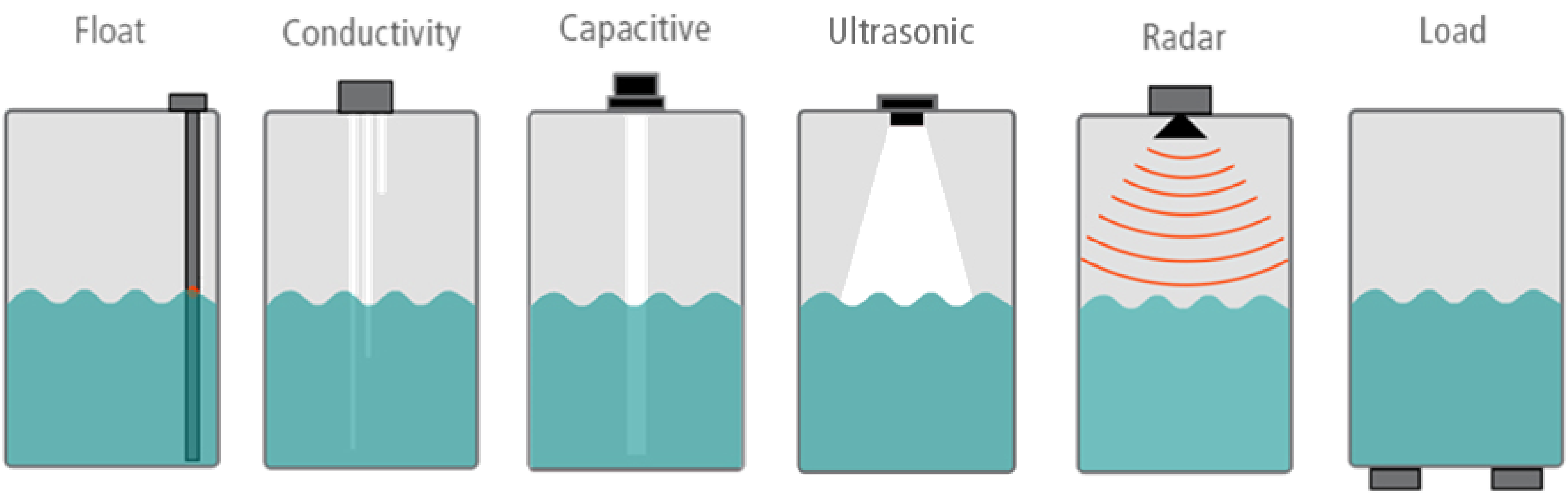
Lens
Focal length
Resolution
Image circle diameter
Optical errors
Mechanical details

Test object
Size
Geometry
Surface
Material
Color

Machine & environment
Cycle time
Feeding system
Required space
Environmental conditions
(Dirt, fog, vibration, daylight)
Interfaces (trigger, digital I/O)

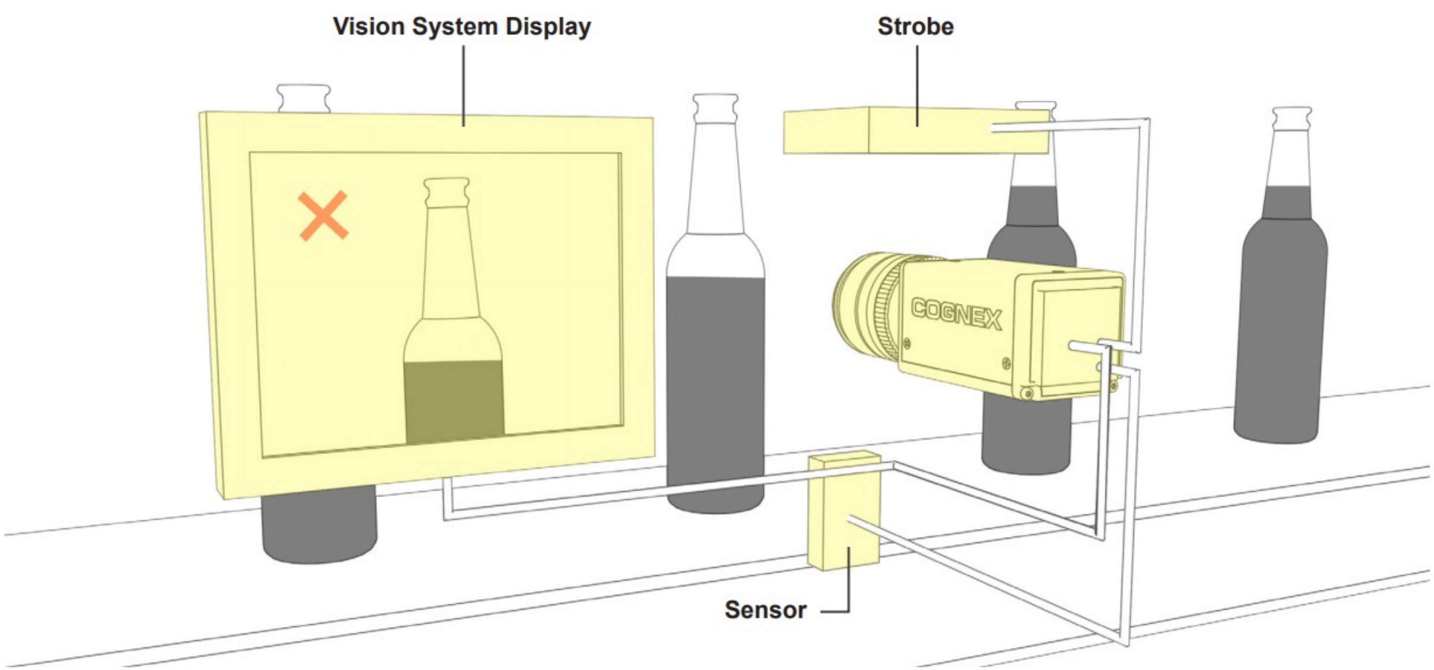
A Fill-level Inspection System

- Design a fill-level inspection system (**without** using the camera)?



A Fill-level Inspection System

- Machine vision-based fill level inspection



- What if the bottle is not transparent?

  X-ray!

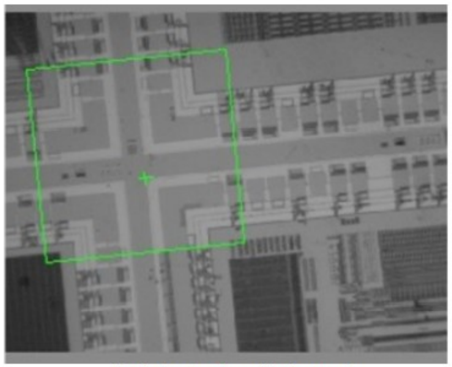
Benefits of Machine Vision

- **Various measurand**: as long as you see it!
- **Measurement speed, accuracy, and repeatability**
  - Fast: hundreds, or even thousands, of parts per minute
  - Detailed inspection: better than human's eye by selecting the right camera resolution and optics
- **Non-contact measurement**
  - Prevents part damage
  - Eliminates the maintenance time and costs
  - Brings safety and operational benefits by reducing human involvement
- **Digital signal**: easy to be processed and automated!

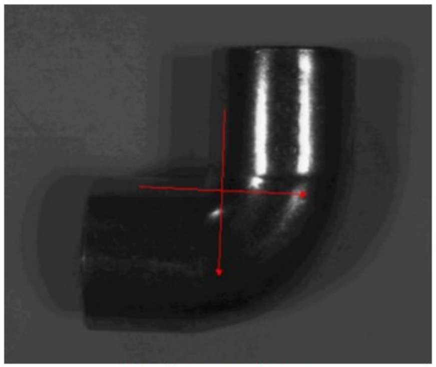
Applications of Machine Vision

- **Guidance**: cooperate with robotic arms

Tomato sauce packets

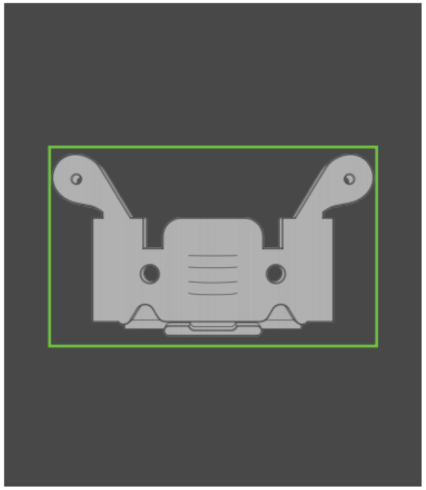


Printed circuit board



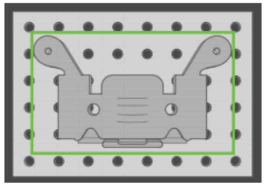
90 degree elbow

Challenging pattern matching scenarios

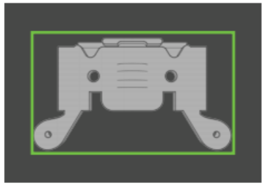


Trained Part

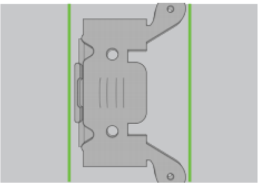
Out of Focus



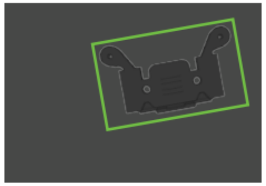
Confusing Background



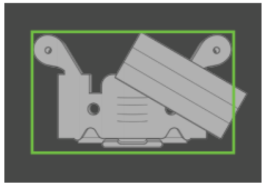
180º Rotation



Reversed Polarity



Scale Change/
Dim Lighting



Occlusion

Applications of Machine Vision

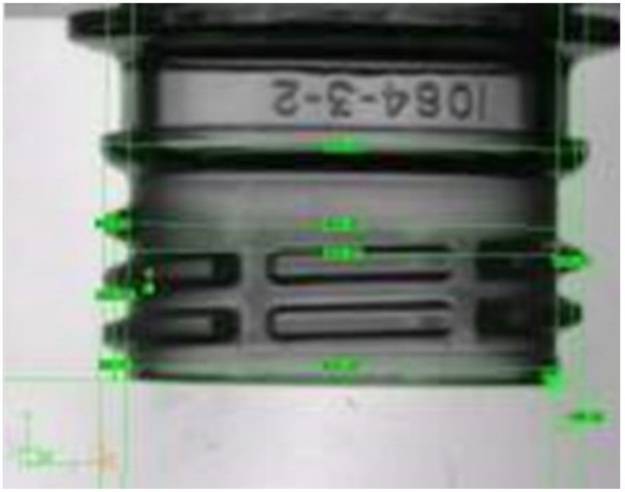
- **Guidance**: blood sample collection

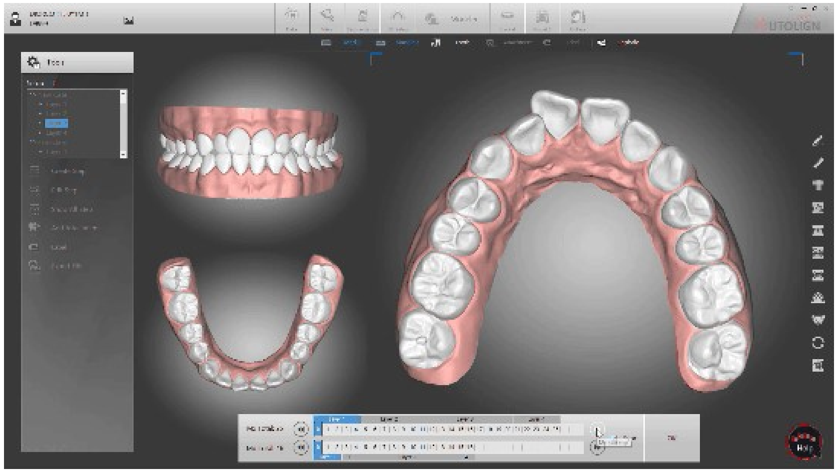
Applications of Machine Vision

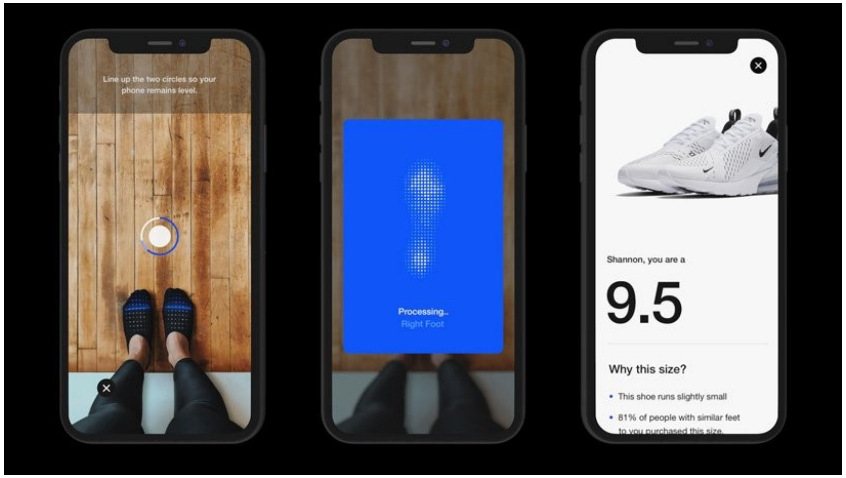
- **Identification**: barcode reading, OCR (optical character recognition)

Applications of Machine Vision

- **Gauging**: 2D/3D distance measurement

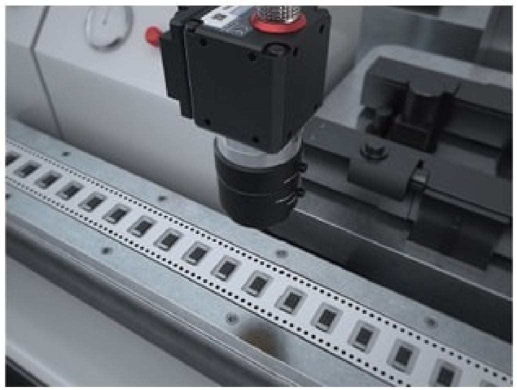




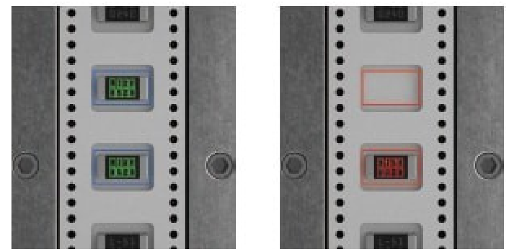


Applications of Machine Vision

- **Inspection**: electronics, automotive, …

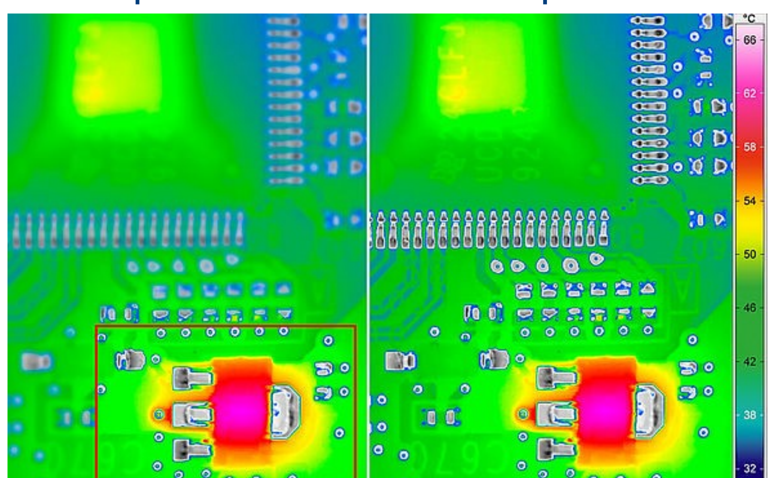


Inspection based on shape



OK image NG image

Inspection based on temperature



Summary: CMOS sensors and camera systems

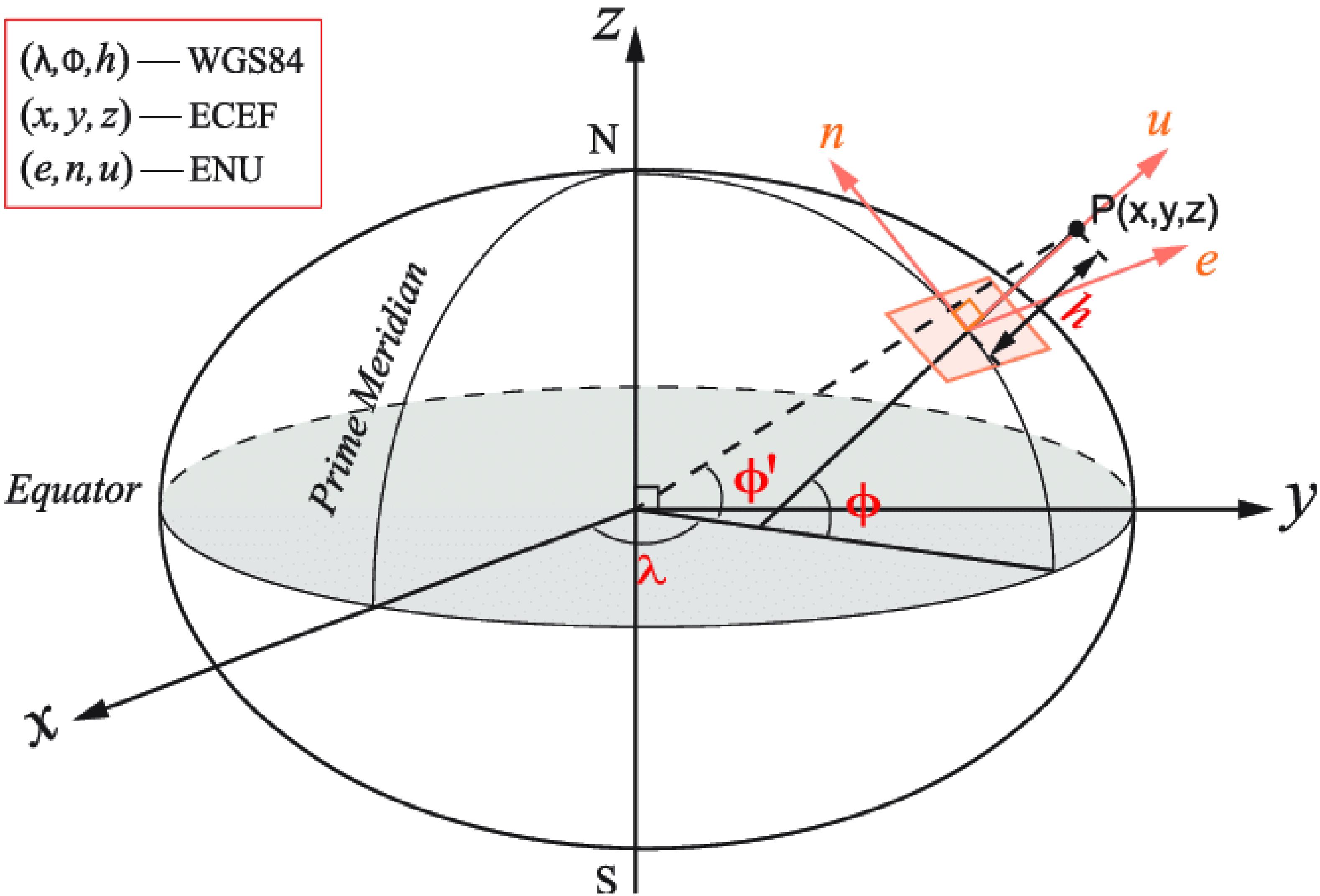
- Principles of image sensors: **Photoelectric Effect**
- Characteristics of image sensors
- Emerging technologies and applications
  - 3-D imaging
  - Event cameras (Dynamic vision sensors)
  - Machine vision systems

Positioning

- Global position
  - Position within **general global reference frame:** longitudes, latitudes, altitudes
- Relative position/Local position
  - Based on **arbitrary coordinate systems and reference frames:** distances between nodes (no relationship to global coordinates)
- Accuracy versus precision
  - e.g., GPS is true within 10m for 90% of all measurements
    Accuracy: 10m (“how close is the reading to the ground truth?”)
    Precision: 90% (“how consistent are the readings?”)

Reference Frames in Positioning

Illustration on the different coordinate systems.

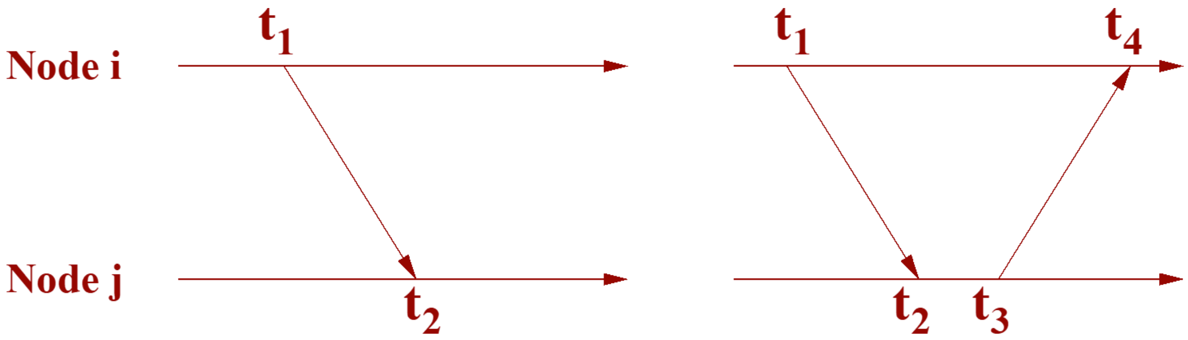


Ranging Techniques

- **Time of Arrival (ToA, time of flight)**
  - determines distance between **sender** and **receiver** of a signal using the measured signal propagation **time** and **known signal velocity:**
    sound waves: 343m/s, i.e., approx. 30ms to travel 10m
    radio signals: 300km/s, i.e., approx. 30ns to travel 10m

- One-way ToA
  - One-way propagation of signal
  - Requires highly accurate synchronization of sender and receiver clocks
- Two-way ToA
  - Round-trip time of signal is measured at sender device
  - Third message if receiver wants to know the distance

Ranging Techniques



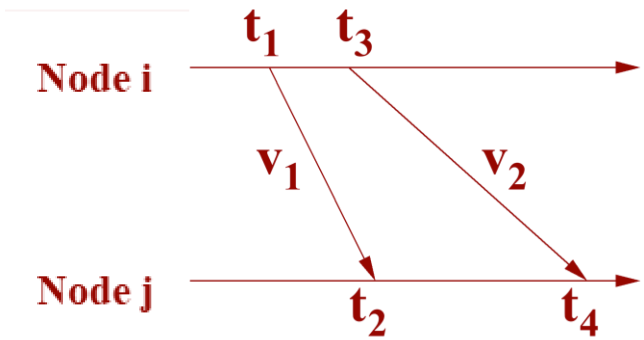
- One-way ToA

$$dist_{ij} = (t_2 - t_1) * v$$

- Two-way ToA

$$dist_{ij} = \frac{(t_4 - t_1) - (t_3 - t_2)}{2} * v$$

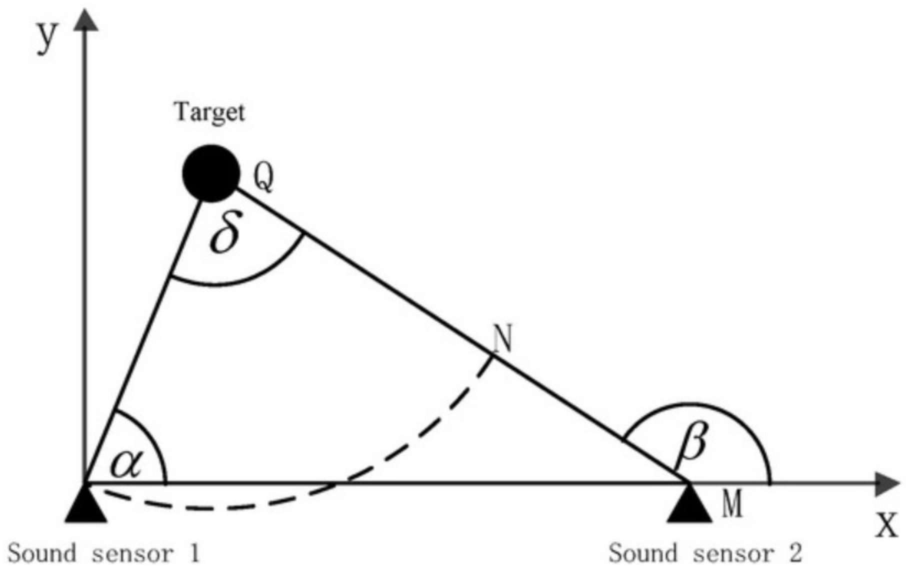
Ranging Techniques



- **Time Difference of Arrival (TDoA)**
  - Two signals with different velocities
  - Example: radio signal (sent at t_1 and received at t_2), followed by acoustic signal (sent at $t_3=t_1+t_{wait}$ and received at t_4)
  - No clock synchronization required
  - Node i as sender only; Node j as receiver only
  - Can you derive the formula to compute the distance?

Ranging Techniques

- **Angle of Arrival (AoA)**
  - Typically achieved using an array of antennas or microphones
  - **Principle:** Spatial separation of antennas or microphones leads to differences in arrival times, amplitudes, and phases
  - Accuracy can be high (within a few degrees)
  - Adds significant hardware cost



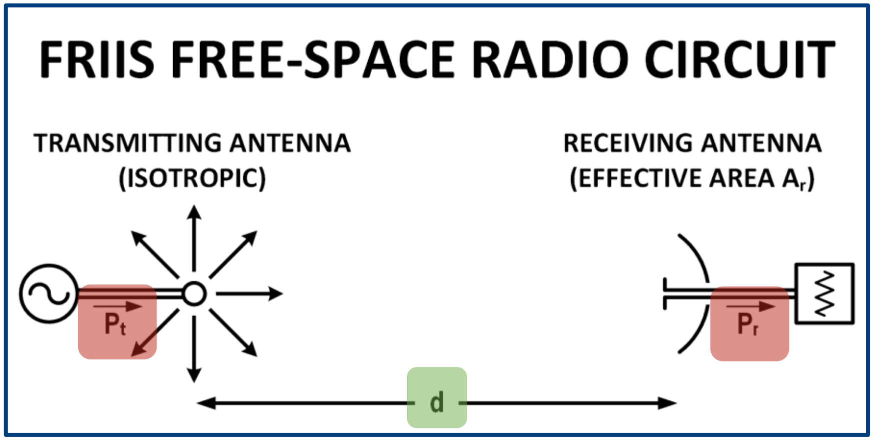
Microphone Array can be used to measure the angle of arrival.

https://www.researcher-app.com/paper/4047874

Ranging Techniques

- **Received Signal Strength (RSS) – Propagation Model**
  - **Principle:** Signal decays with distance
  - Many devices measure signal strength with Received Signal Strength Indicator (RSSI):
    Vendor-specific interpretation and representation
  - In free space, **RSS degrades with square of distance**, expressed by Friis transmission equation

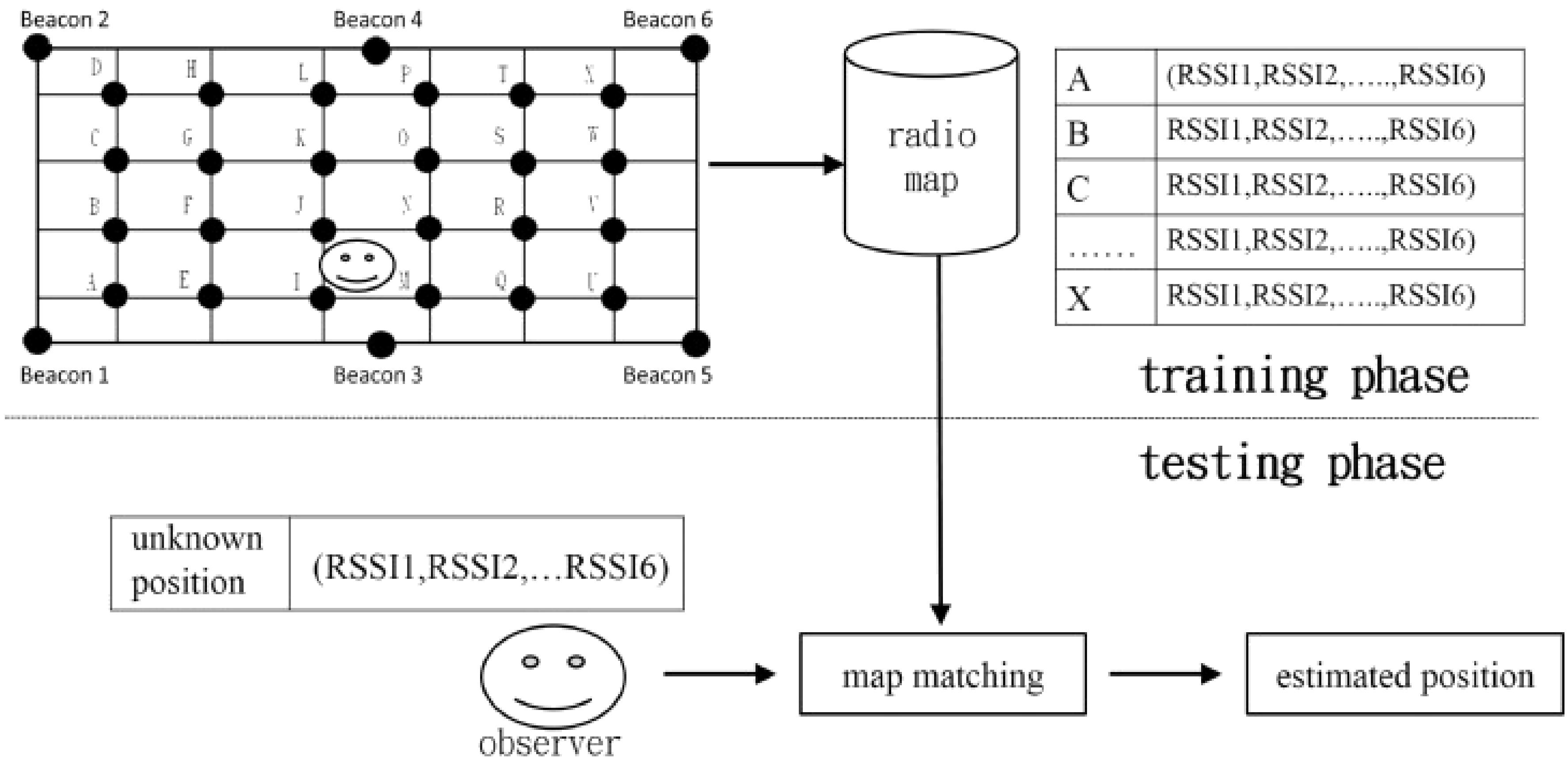
$$\frac{P_r}{P_t} = D_t D_r \left(\frac{\lambda}{4\pi d}\right)^2$$



https://en.wikipedia.org/wiki/Friis_transmission_equation

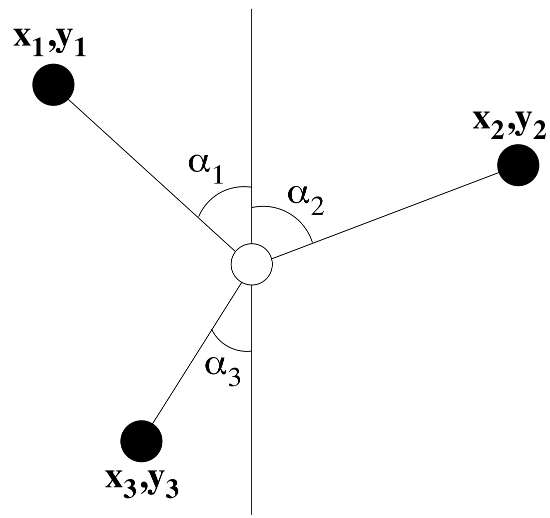
Fingerprint Matching

- Friis transmission equation may have large errors in practice due to obstacles, multi-path effects, and inferences…
- Fingerprint Matching



Triangulation and Trilateration

- **Triangulation:** With the angle, how to find the position?
- Uses the geometric properties of triangles to estimate location
- Minimum of **2** bearing lines (and the locations of anchor nodes or the distance between them) are needed for two-dimensional space

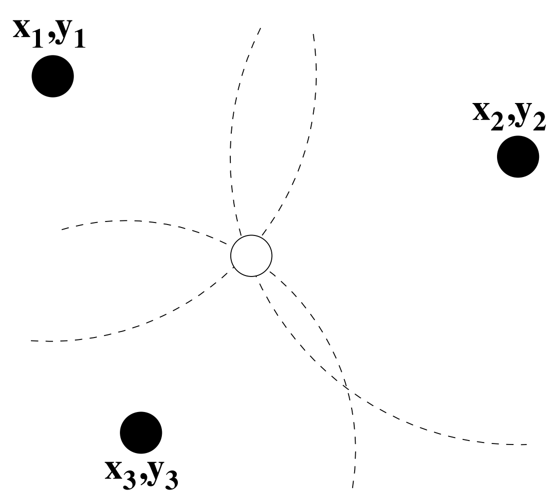


Triangulation and Trilateration

- **Trilateration:** With the range, how to find the position?
- Idea: given the distance to an anchor, it is known that the node must be along the circumference of a circle centered at anchor and a radius equal to the node-anchor distance.
- In 2-D space, how many non-collinear anchor required?

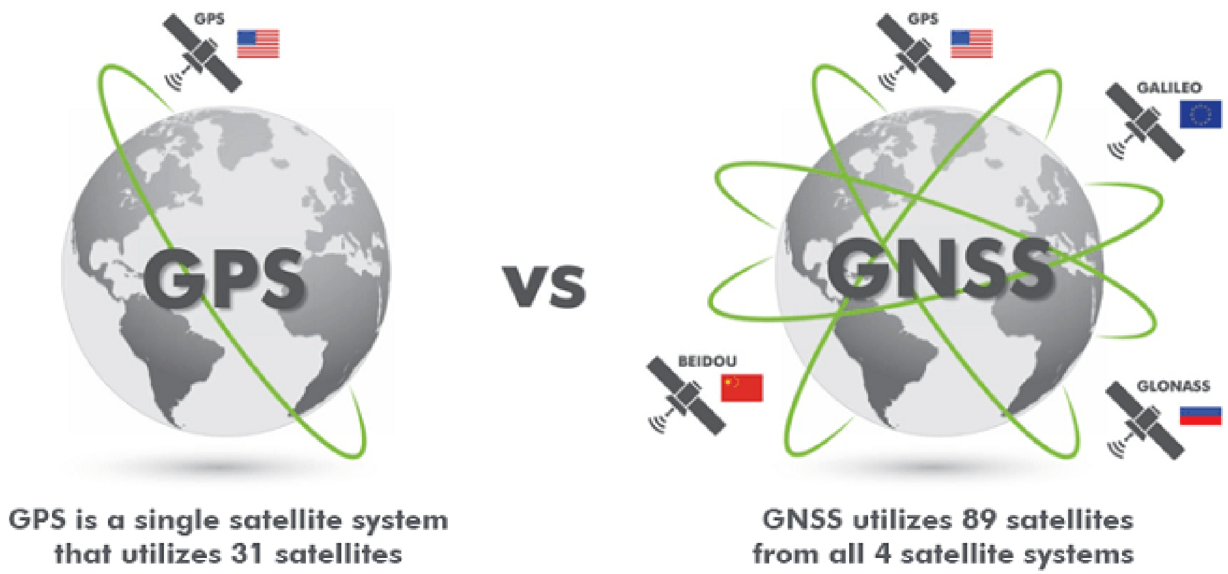
  At least 3.
- In 3-D space, how many non-coplanar anchor required?

  At least 4.



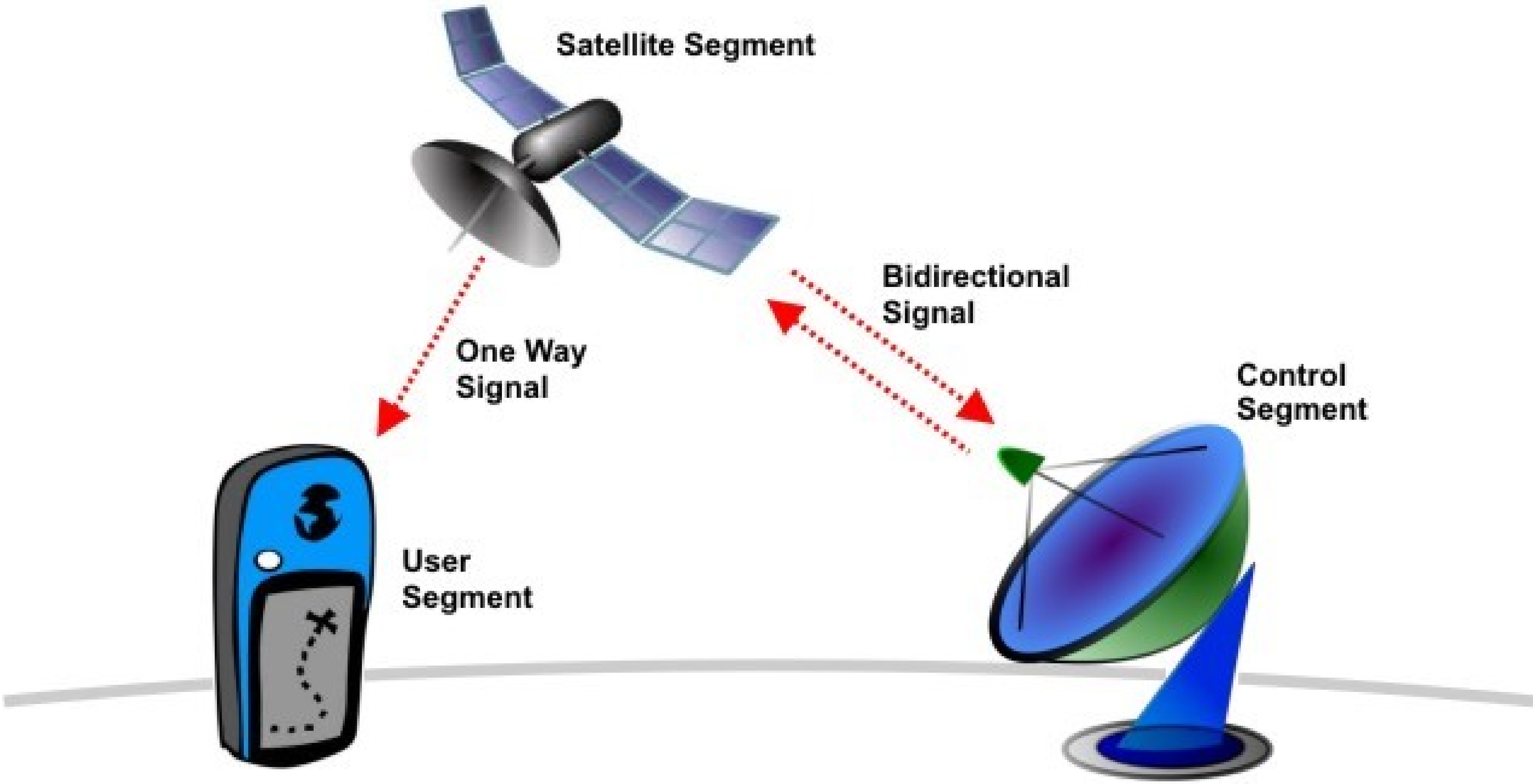
GNSS (Global Navigation Satellite System)

- GPS is just one of (but maybe the most popular) GNSSs.
  - Started in 1973, fully operational in 1995
- GPS two levels of service
  - Standard Positioning Service (SPS): available to all users, high-quality receivers have accuracies of up to 100m (before 2010), 3m (after 2010)
  - Precise Positioning Service (PPS): used by US and Allied military users



GNSS (Global Navigation Satellite System)

- Communications and connections in GNSS

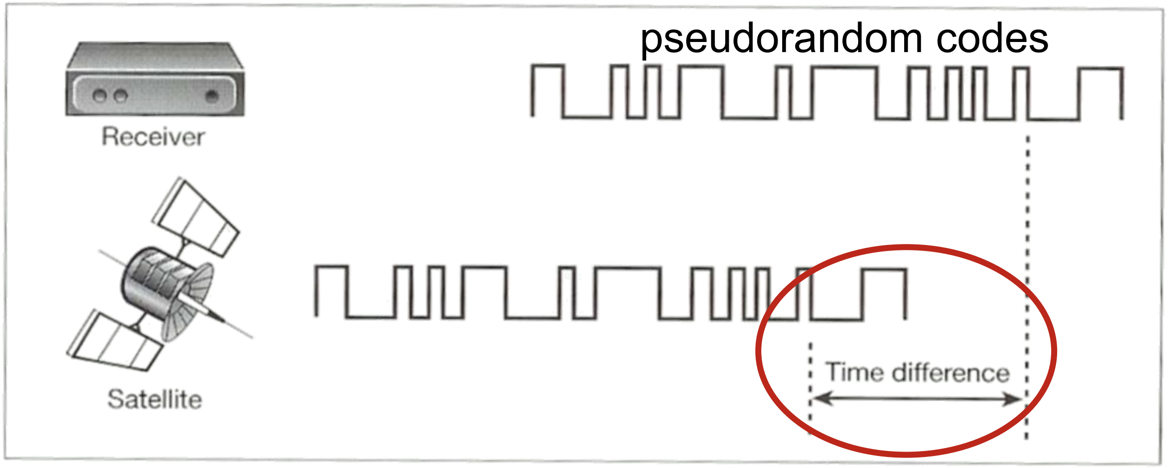


GNSS Principle

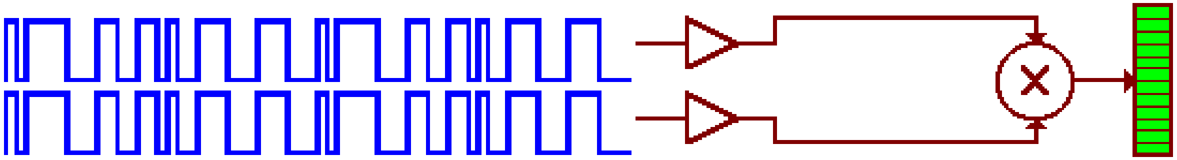
- Satellites and receivers use accurate and synchronized clocks
- Receiver compares **generated code** with **received code** to determine
  - The time difference Δ between code generation time and current time

  → Δ expresses the **travel time of the code** from satellite to receiver

Time lag between pseudorandom codes generated by the satellite and the receiver is used to measure distance.



Use the **correlation** to measure the time lag!

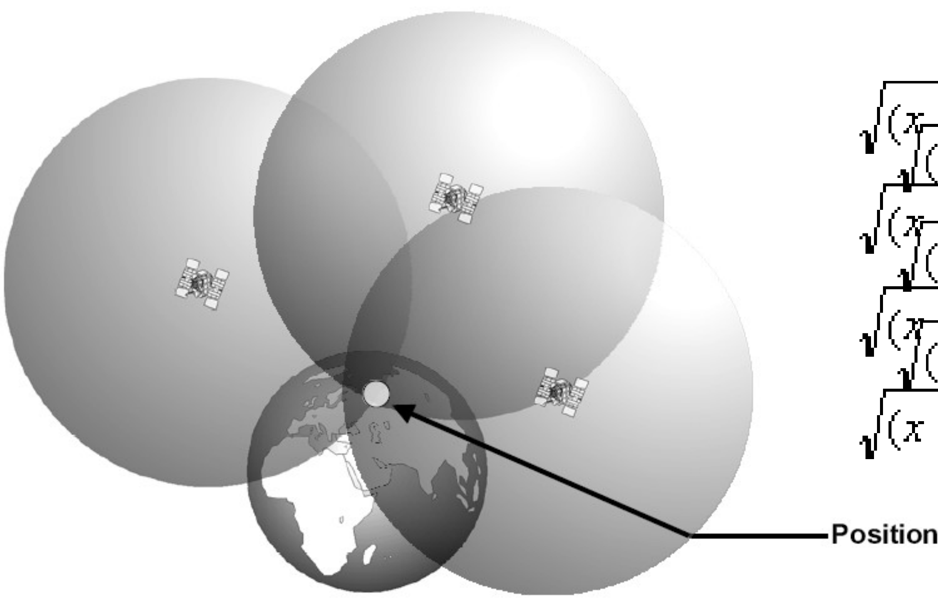


GNSS Principle

- We already know the travel time from satellite to receiver.
  How to obtain the distance?

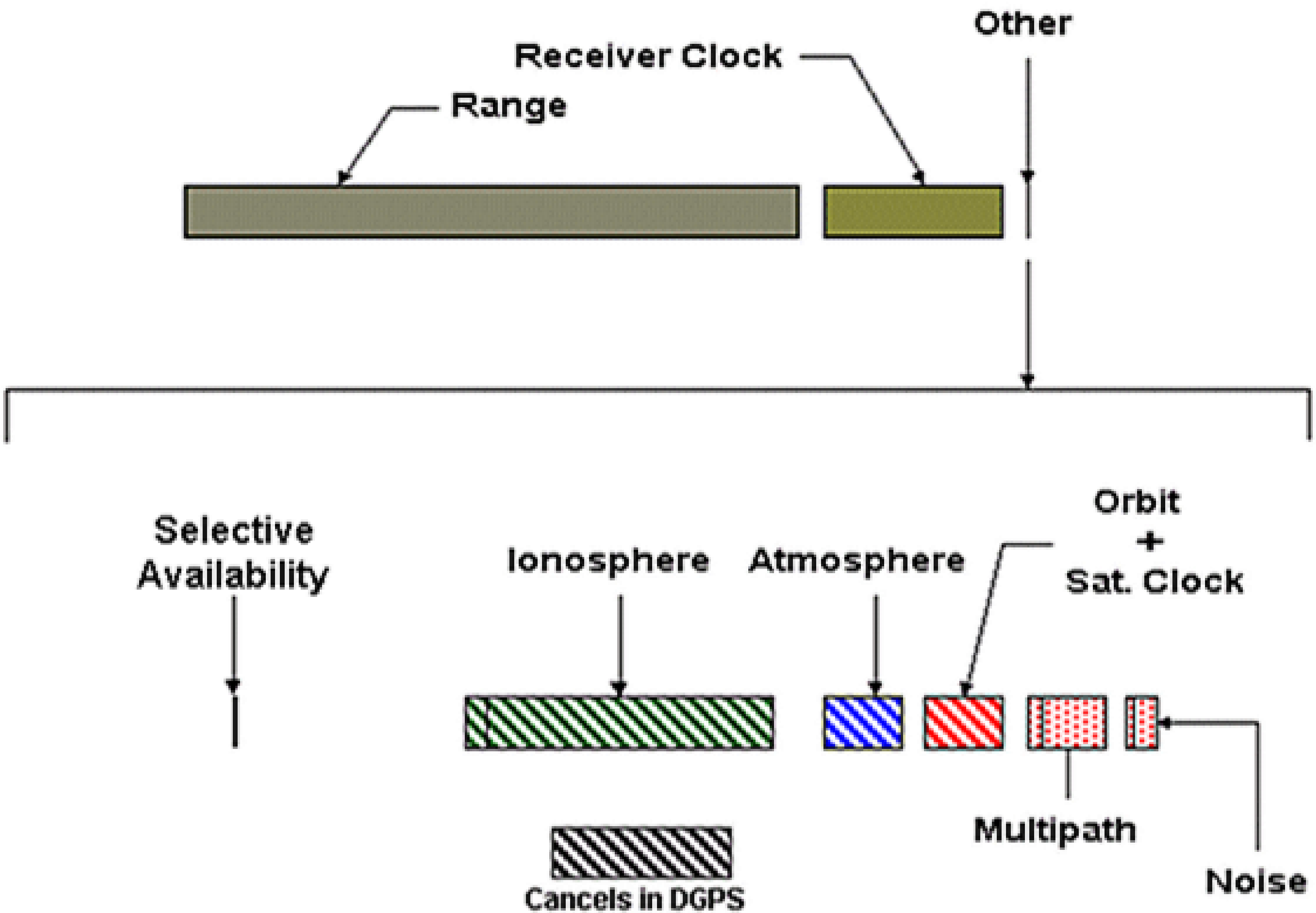
Distance = time * velocity

- How many satellites required to solve GPS receiver position?



$$\sqrt{(x-x_1)^2+(y-y_1)^2+(z-z_1)^2}+ct_B=d_1$$
$$\sqrt{(x-x_1)^2+(y-y_1)^2+(z-z_1)^2}=d_1$$
$$\sqrt{(x-x_2)^2+(y-y_2)^2+(z-z_2)^2}+ct_B=d_2$$
$$\sqrt{(x-x_2)^2+(y-y_2)^2+(z-z_2)^2}=d_2$$
$$\sqrt{(x-x_3)^2+(y-y_3)^2+(z-z_3)^2}+ct_B=d_3$$
$$\sqrt{(x-x_3)^2+(y-y_3)^2+(z-z_3)^2}=d_3$$
$$\sqrt{(x-x_4)^2+(y-y_4)^2+(z-z_4)^2}+ct_B=d_4$$

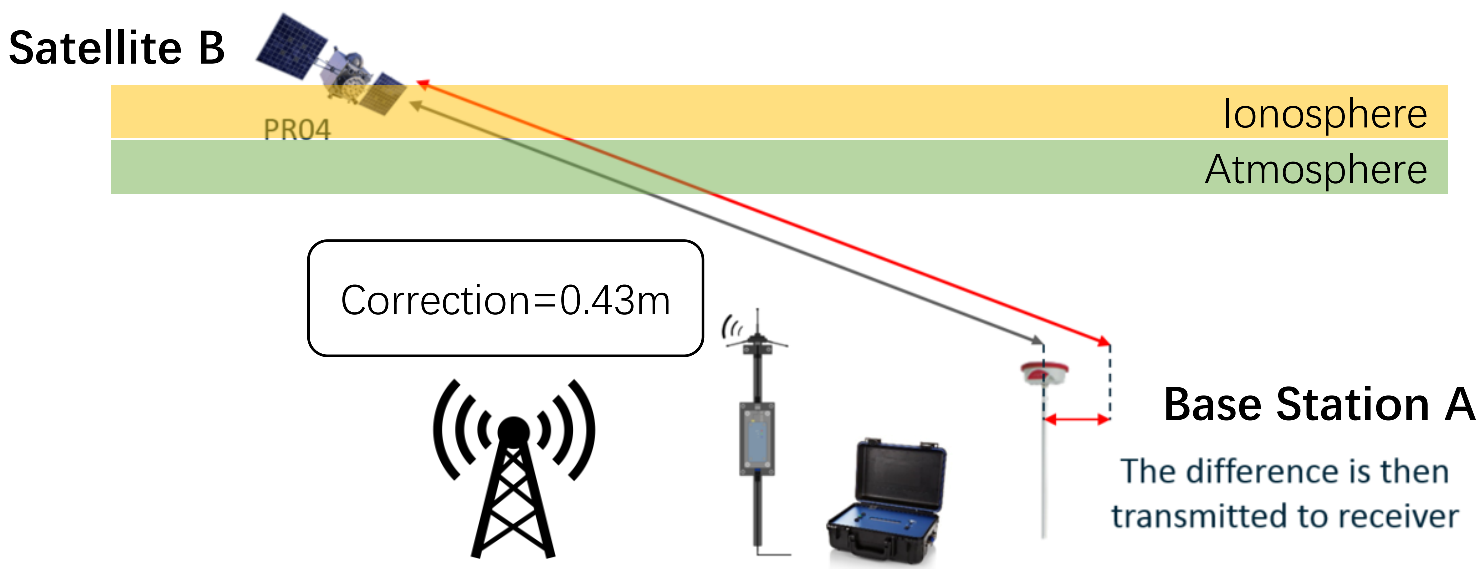
GNSS Principle: Errors



Improving GNSS Accuracy: Differential GPS

- Idea: Add one **Base Station.** we already know:
  - The exact position A of the base station on the Earth
  - The exact position B of the Satellite

  Real distance between A and B
- And we have the "**measured distance**" between A and B
- By differencing real distance and measured distance, we compute the correction! Correction is then broadcasting to users' receivers.



Improving GNSS Accuracy: Differential GPS

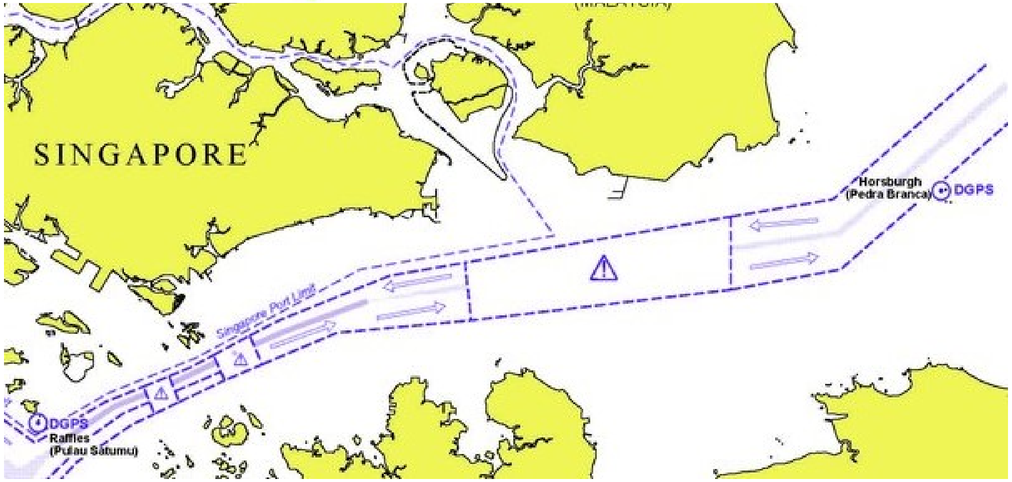
- D-GPS Accuracy (under optimal conditions)
  - User hand held: 2-5 m
  - Survey grade units: < 0.03 m
  - Very high precision units: ~ 0.005 m

Improving GNSS Accuracy: Differential GPS

- D-GPS in Singapore

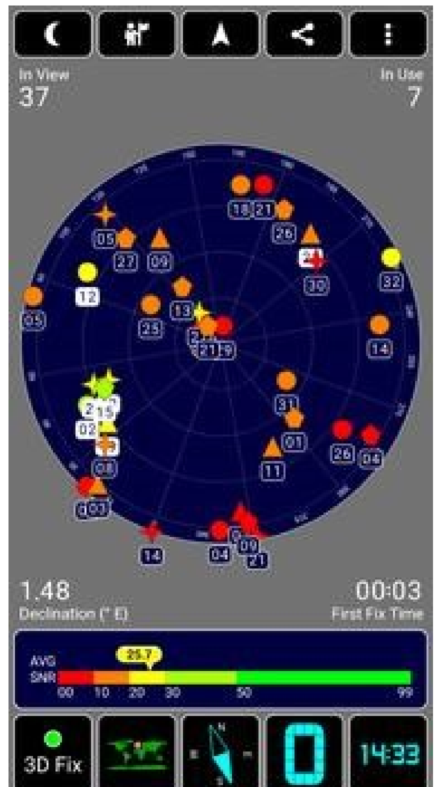
SLA
SINGAPORE
LAND AUTHORITY

MPA
SINGAPORE



GPS Test App on Smartphone

- Check GPS signal and status.
- Viewing: your altitude, accurate UTC plus local time, your compass heading, the moon phase and daylight hours.



Wi-Fi Localization

- GNSS is good. What about indoors?
- Wi-Fi access points are everywhere, and broadcast signals up to 100m
- Wi-Fi chips in devices detect the **name of the access point**, **signal strength**, and (sometimes) angle of arrival

  - Model-based localization
  - **Learning-based localization**

Learning-based Wi-Fi Localization

- Training: Create fingerprints

AP1: -75 dBm
AP2: -52 dBm
AP3: -80 dBm

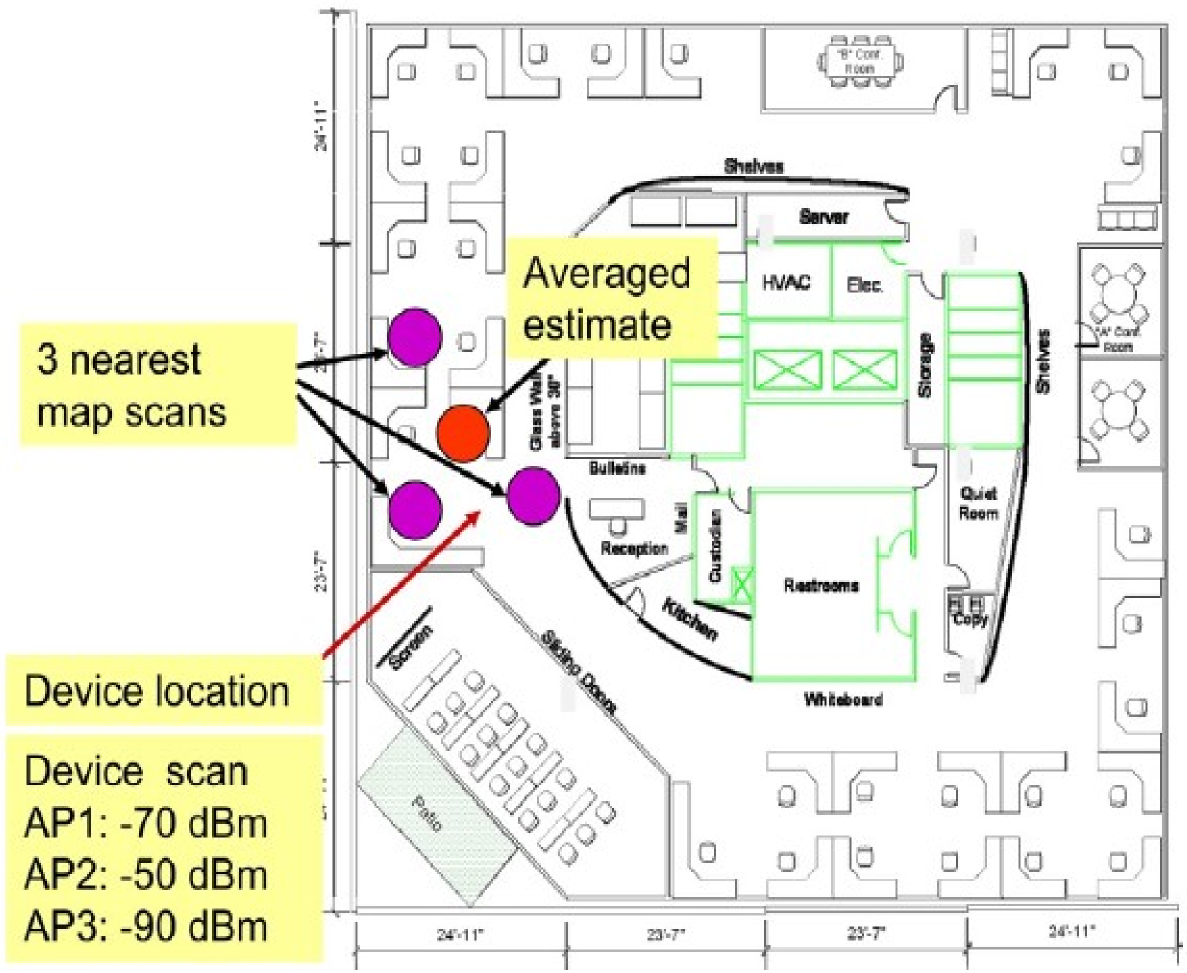
AP1: -45 dBm
AP2: -58 dBm
AP3: -85 dBm

AP4: -35 dBm
AP5: -64 dBm
AP6: -72 dBm

dBm: decibel-milliwatts, a unit to express the power level with reference to 1 mW.

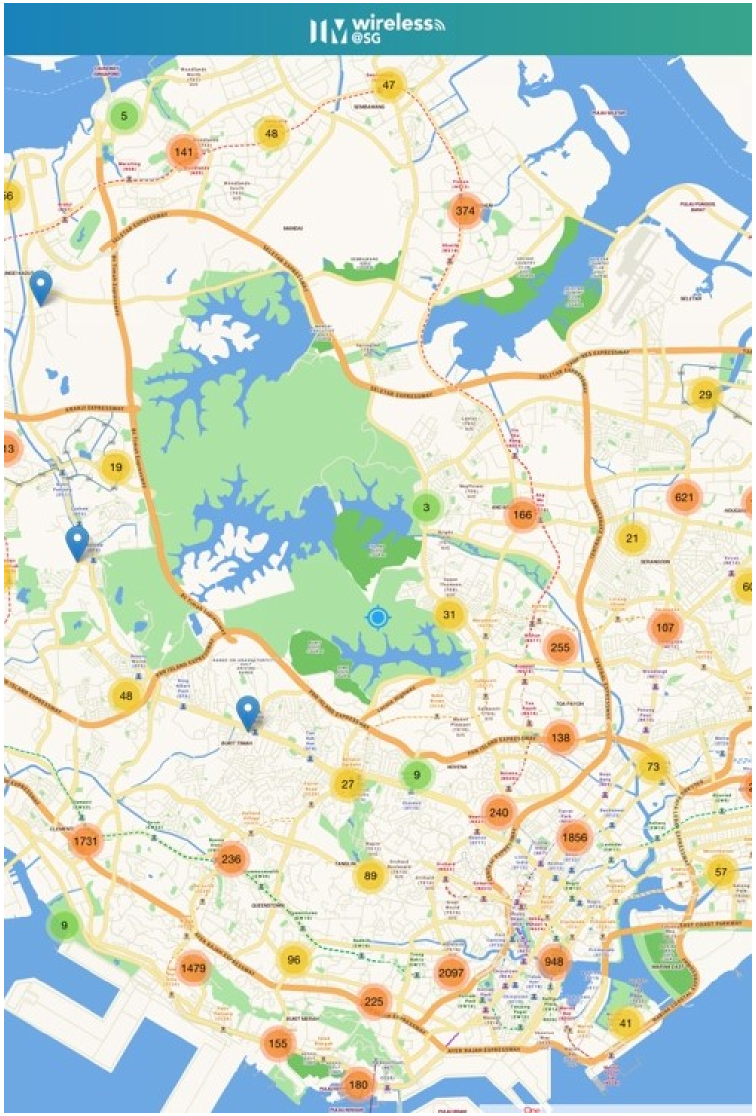
Learning-based Wi-Fi Localization

- Testing: Find the most possible matching



Wi-Fi Localization: Summary

- Wi-Fi is everywhere now
  - No new infrastructure
  - Low cost
- Position using Wi-Fi
  - Gives **2-3 m** indoor accuracy
  - Requires high calibration overhead: 10+ hours per building
  - Changes over time (adding/removing/relocating APs) impact accuracy
  - Requires maintenance on AP maps



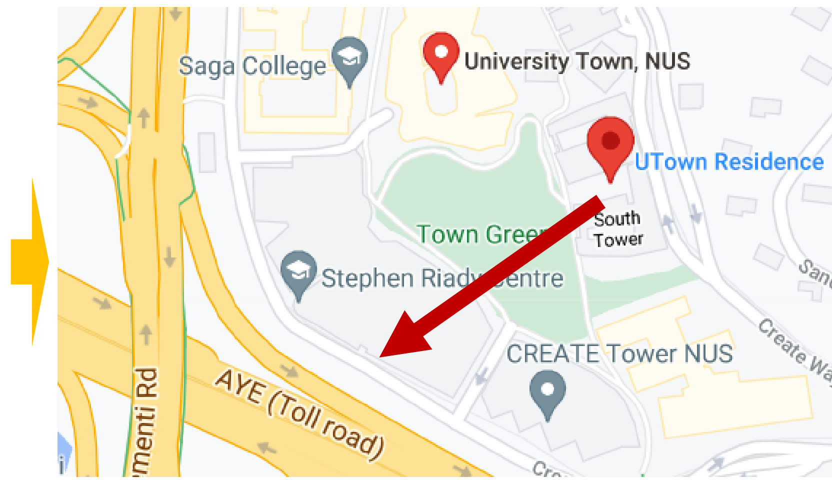
Video: Other Indoor Positioning Techniques

Summary: Positioning systems

- Reference frames
- Ranging Techniques
  - Time of Arrival
  - Time Difference of Arrival
  - Angle of Arrival
  - Received Signal Strength (RSS) – propagation model
  - Fingerprint matching
- Triangulation and Trilateration
- Common positioning techniques
  - GNSS-based, WiFi-based

Vision-based Pose Estimation

- Human can estimate the location and pose using eyes.
  Two types of vision-based location/pose estimation will be discussed:
  - **Absolute location/pose:**



Visual place recognition

Vision-based Pose Estimation

- **Relative location/pose:**

What is the translation/rotation between two frames?

Visual odometry (VO)

[The KITTI Vision Benchmark Suite - YouTube]

Visual Place Recognition: Challenges

- The appearance of a place can change drastically
  - (a) same location with different appearance
  - (b) different locations with similar appearance

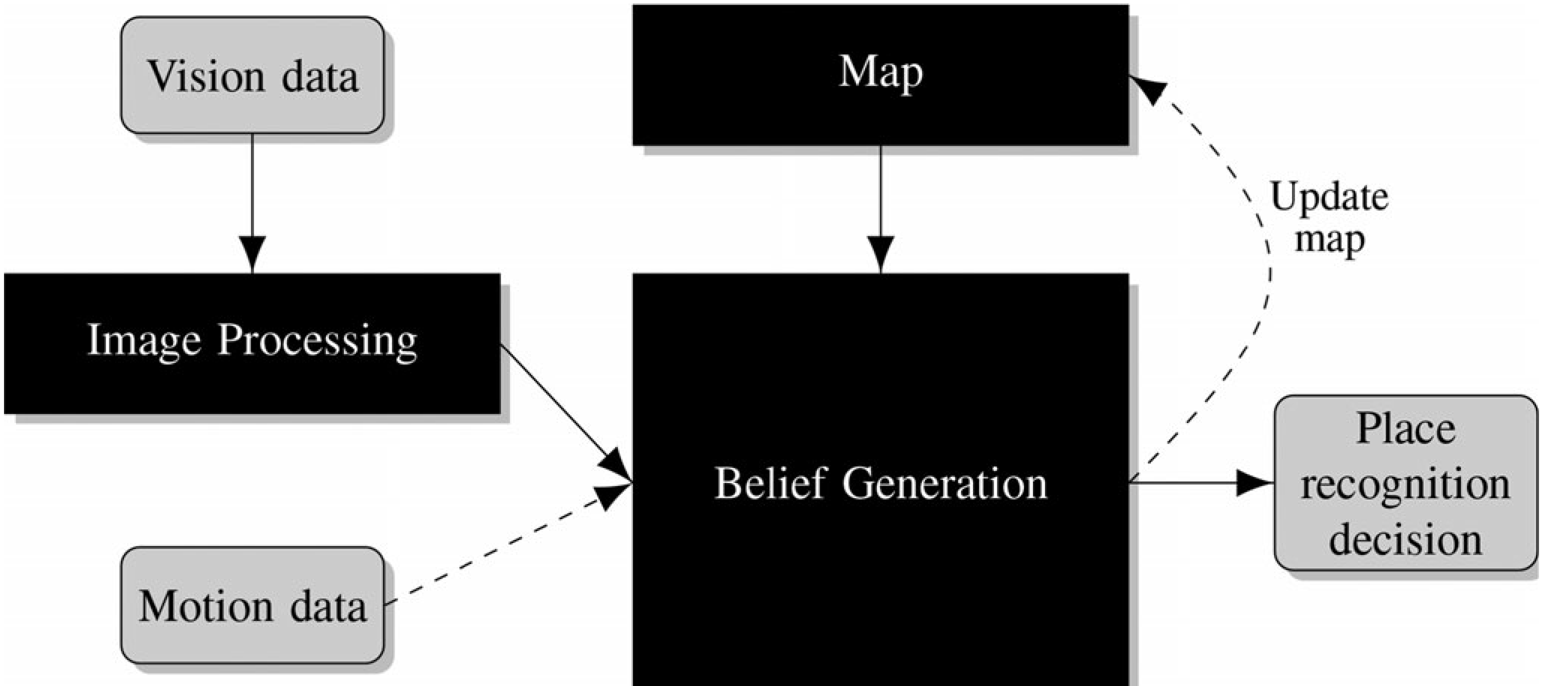

(a)


(b)

intro.pdf - Google Drive
Visual Localization - ECCV 2018 Tutorial (google.com)
7 - visual place recognition in changing environments · Visual Place Recognition: A Survey (gitbooks.io)

Visual Place Recognition System

stores robot's
knowledge of the world



optional

decides whether the current visual data
matches a previously stored place

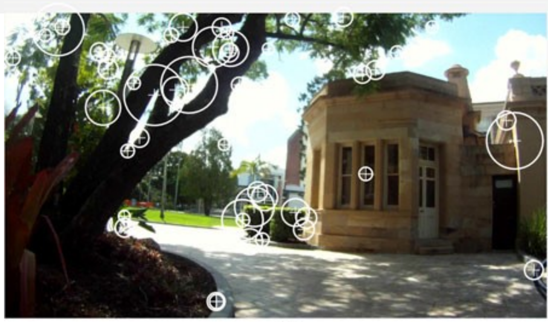
[Intro.pdf - Google Drive](#)
[Visual Localization - ECCV 2018 Tutorial (google.com)](#)
[7 - visual place recognition in changing environments · Visual Place Recognition: A Survey (gitbooks.io)](#)

The Image Processing Module

- How to describe places?
  - local feature descriptor, right fig. (a)
  - global / whole-image descriptors, right fig. (b)
- **Local feature descriptors**
  1. Detect the salient parts (“features”) of the image.
  2. Store the contents of features and nearby pixels.

     detection → description

  3. To increase efficiency, convert the extracted features into a vocabulary that can be compared using text retrieval techniques.
     e.g., using the bag-of-words model.



(a)

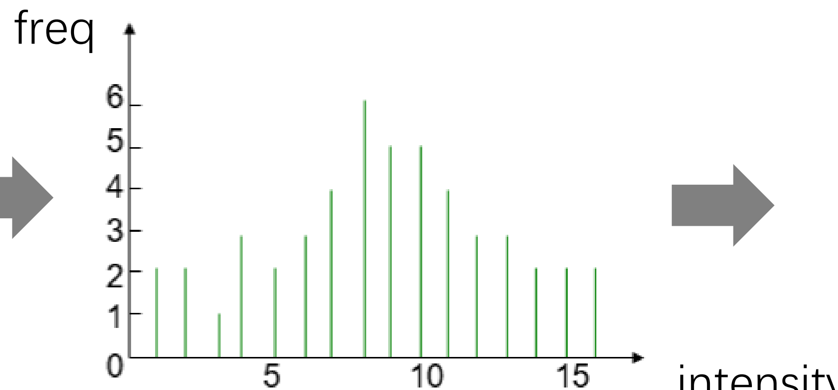
(b)

[Intro.pdf - Google Drive](#)
[Visual Localization - ECCV 2018 Tutorial (google.com)](#)
[7 - visual place recognition in changing environments · Visual Place Recognition: A Survey (gitbooks.io)](#)

The Image Processing Module

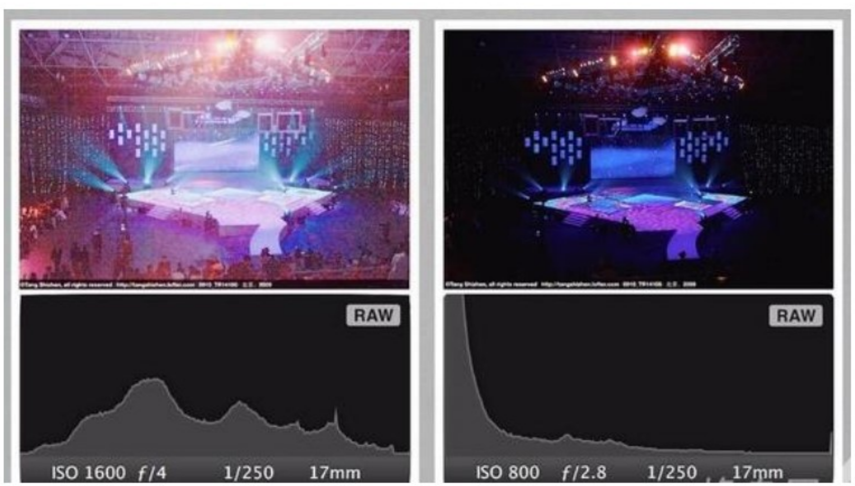
- **Global / whole-image descriptors:**
  - e.g., Histogram

1	7	8	9	10	11	14
5	2	6	7	14	12	15
3	4	7	8	6	9	11
2	1	4	7	8	8	9
8	4	5	9	11	12	10
8	10	11	15	16	10	13
13	6	9	16	13	12	10



vector representation

2
2
1
…



Intensity histogram under different light conditions.

《数字图像分析》课程提纲 (ustc-dia.github.io)
7 - visual place recognition in changing environments · Visual Place Recognition: A Survey (gitbooks.io)

The Mapping Module

- How to remember places?
  - Different map definitions and representations.

MAPPING FRAMEWORKS FOR VISUAL PLACE RECOGNITION

Level of map abstraction	Place description type	Comments
Pure image retrieval	Appearance-based	No position information
Topological	Appearance-based	Includes transition information
Topological-metric	Appearance-based	Includes metric information between but not within places
Topological-metric	Sparse metric information (landmark maps)	SLAM system – includes metric information between and within places
Topological-metric	Dense metric information (occupancy grid maps)	SLAM system – includes metric information between and within places

Intro.pdf - Google Drive
Visual Localization - ECCV 2018 Tutorial (google.com)
7 - visual place recognition in changing environments · Visual Place Recognition: A Survey (gitbooks.io)

The Belief Generation Module

- How to recognize places?
  - **Naïve matching**: compute the similarity of the query image and database images, then select the most probable ones. For BoW model, the similarity is higher if:
    more features exists in both query image and database images.
  - **Combine** the map, transition, and visual appearance:
    → SLAM problem:
    filtering-based approach / optimization-based approach

[Intro.pdf - Google Drive](#)
[Visual Localization - ECCV 2018 Tutorial (google.com)](#)
[7 - visual place recognition in changing environments · Visual Place Recognition: A Survey (gitbooks.io)](#)

Visual Odometry (VO)

- VO Working Principle

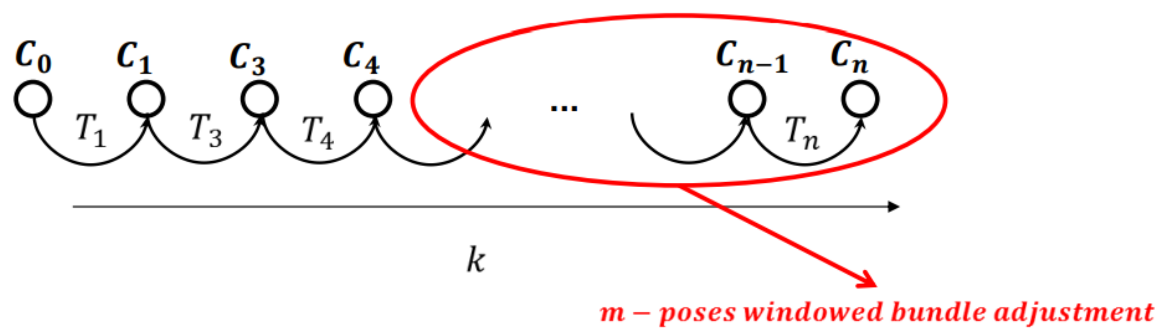
1. Compute the relative motion T_k from images I_{k-1} to image I_k

$$T_k = \begin{bmatrix} R_{k,k-1} & t_{k,k-1} \\ 0 & 1 \end{bmatrix}$$

2. Concatenate them to recover the full trajectory

camera pose $\quad C_n = C_{n-1}T_n \quad$ transformation

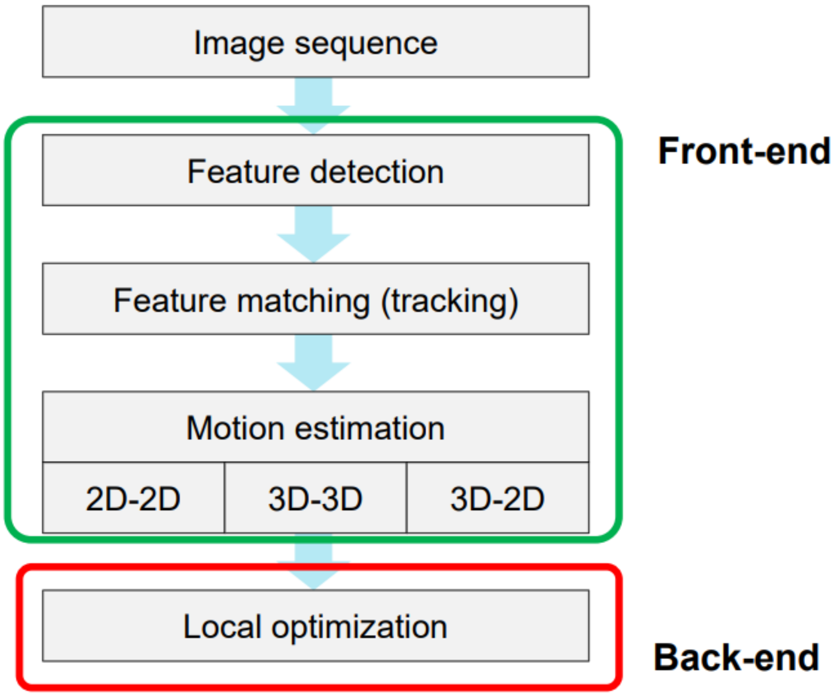
3. An optimization over the last m poses can be done to refine locally the trajectory (Pose-Graph or Bundle Adjustment)



[(5) (PDF) Visual Odometry [Tutorial] (researchgate.net)]

Front-End and Back-End

- VO systems contains the front-end and the back-end.
- The **front-end** is responsible for
  - Feature extraction, matching, and outlier removal
  - Loop closure detection
  - Motion estimation between two frames
- The **back-end** is responsible for the pose and structure optimization (e.g., with libraries iSAM, g2o, Google Ceres)

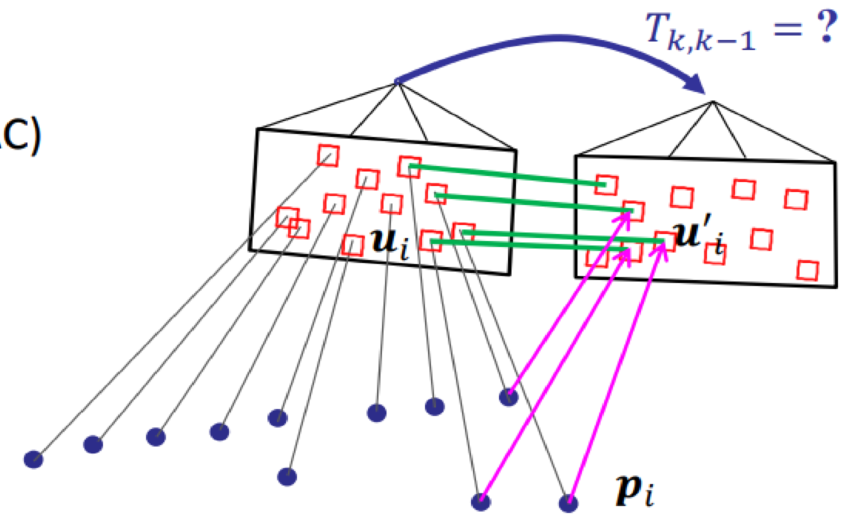


(5) (PDF) Visual Odometry [Tutorial] (researchgate.net)

Estimating the Relative Motion T_k

Feature-based methods

1. Extract & match features (+RANSAC)
2. Minimize **Reprojection error** minimization



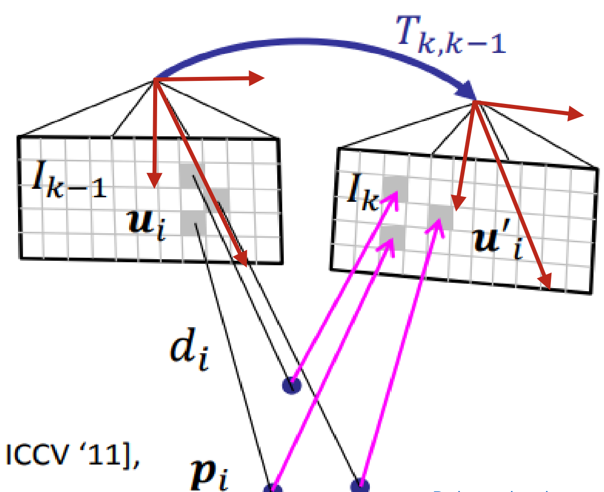
$$T_{k,k-1} = \arg\min_{T} \sum_i \|\boldsymbol{u'}_i - \pi(\boldsymbol{p}_i)\|^2_{\Sigma}$$

Direct methods

1. Minimize **photometric error**

$$T_{k,k-1} = \arg\min_{T} \sum_i \|I_k(\boldsymbol{u'}_i) - I_{k-1}(\boldsymbol{u}_i)\|^2_{\sigma}$$

where $\boldsymbol{u'}_i = \pi\left(T \cdot (\pi^{-1}(\boldsymbol{u}_i) \cdot d)\right)$



[Jin,Favaro,Soatto'03] [Silveira, Malis, Rives, TRO'08], [Newcombe et al., ICCV '11], [Engel et al., ECCV'14], [Forster et al., ICRA'14]

Robots that know what they do (upenn.edu)

Estimating the Relative Motion T_k

Feature-based methods

1. Extract & match features (+RANSAC)
2. Minimize **Reprojection error** minimization

$$T_{k,k-1} = \arg\min_T \sum_i \| \boldsymbol{u}'_i - \pi(\boldsymbol{p}_i) \|_\Sigma^2$$

✓ Large frame-to-frame motions

✓ Accuracy: Efficient optimization of structure and motion (Bundle Adjustment)

× Slow due to costly feature extraction and matching

× Matching Outliers (RANSAC)

Direct methods

1. Minimize **photometric error**

$$T_{k,k-1} = \arg\min_T \sum_i \| I_k(\boldsymbol{u}'_i) - I_{k-1}(\boldsymbol{u}_i) \|_\sigma^2$$

$$\text{where} \quad \boldsymbol{u}'_i = \pi\left(T \cdot (\pi^{-1}(\boldsymbol{u}_i) \cdot d)\right)$$

✓ All information in the image can be exploited (precision, robustness)

✓ No need to calculate feature points and descriptors.

× Limited frame-to-frame motion

× Joint optimization of dense structure and motion too expensive

[Jin,Favaro,Soatto'03] [Silveira, Malis, Rives, TRO'08], [Newcom
[Engel et al., ECCV'14], [Forster et al., ICRA'14]

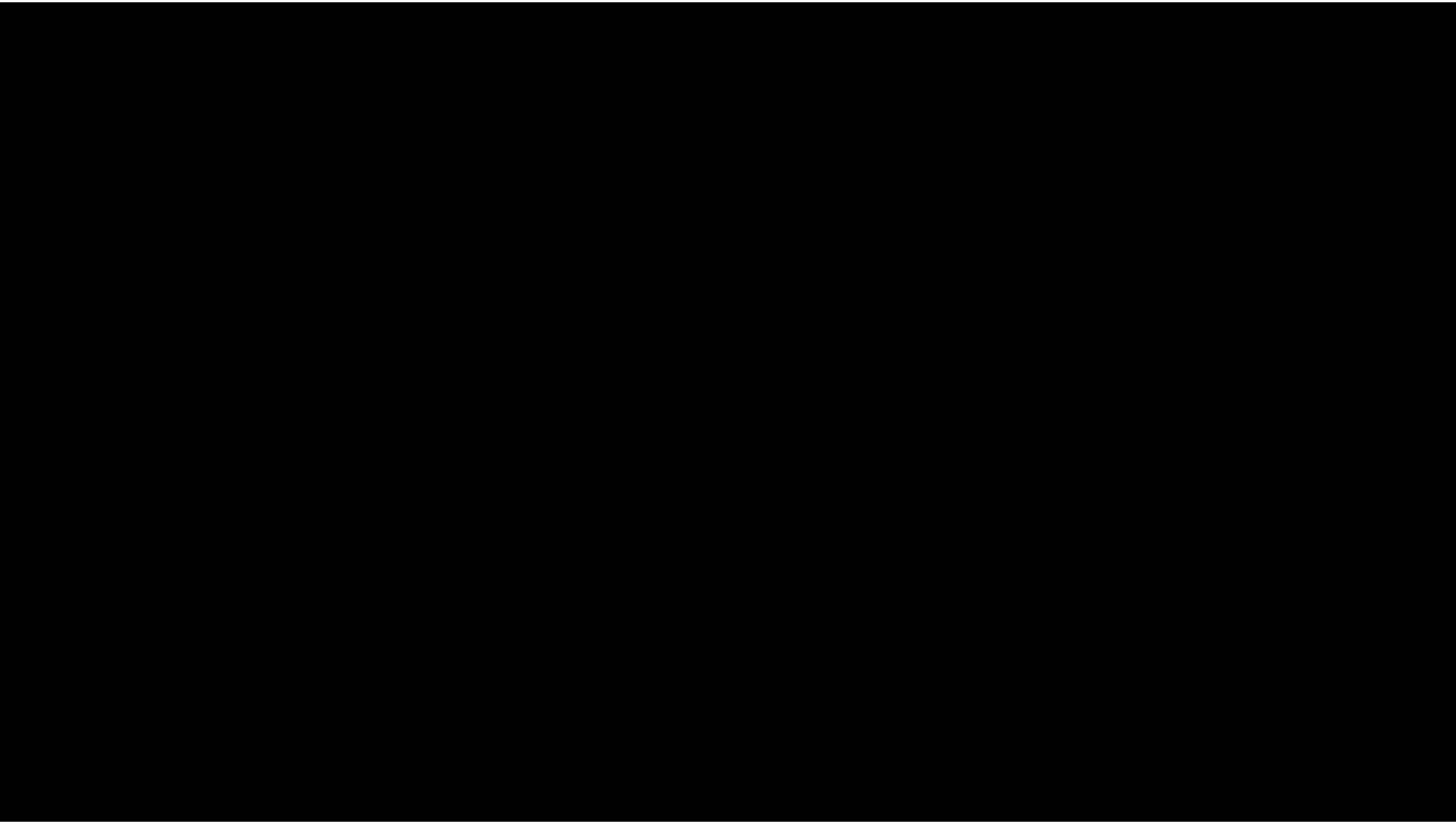
Robots that know what they do (upenn.edu)

VO Challenges

- Sufficient illumination in the environment
- Dominance of static scene over moving objects
- Enough texture to allow apparent motion to be extracted
- Sufficient scene overlap between consecutive frames

Robots that know what they do (upenn.edu)

KITTI Dataset for Algorithm Evaluation



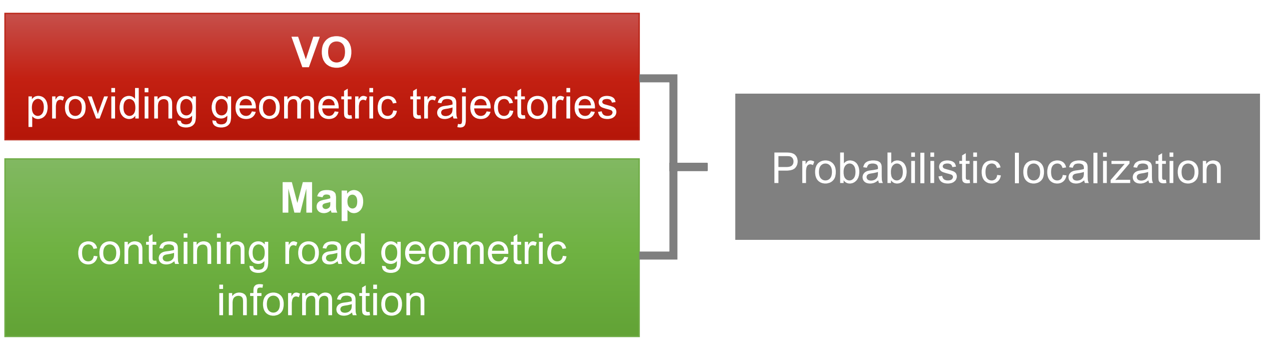
The KITTI Vision Benchmark Suite (cvlibs.net)

Visual Odometry + Road Maps = Self-Localization

- Limitations of GPS and place recognitions
  - GPS has limitations in terms of reliability and availability.
  - Place recognition techniques use image features and a database of previously collected images.

What if we do not have the database for place recognition, and GPS is not available?

- This work develop an inexpensive technique for localizing to ~3m in **unseen regions. How?**



Assumption: vehicle is on the road at all times.

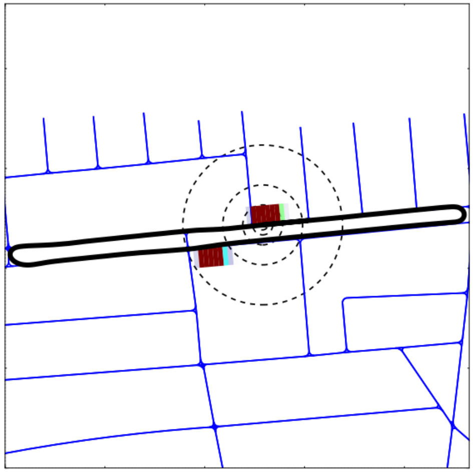
CVPR13.pdf (toronto.edu)
CVPR13-Slides.pdf (toronto.edu)
Lost! Leveraging the Crowd for Probabilistic Visual Self-Localization (toronto.edu)

Visual Odometry + Road Maps = Self-Localization

- Experiments: Quantitative Results

Average	Stereo Odometry	Monocular Odometry	Map Projection
Position Error	3.1m	18.4m	1.4m
Heading Error	1.3°	3.6°	-
Localization Time	36s	62s	-

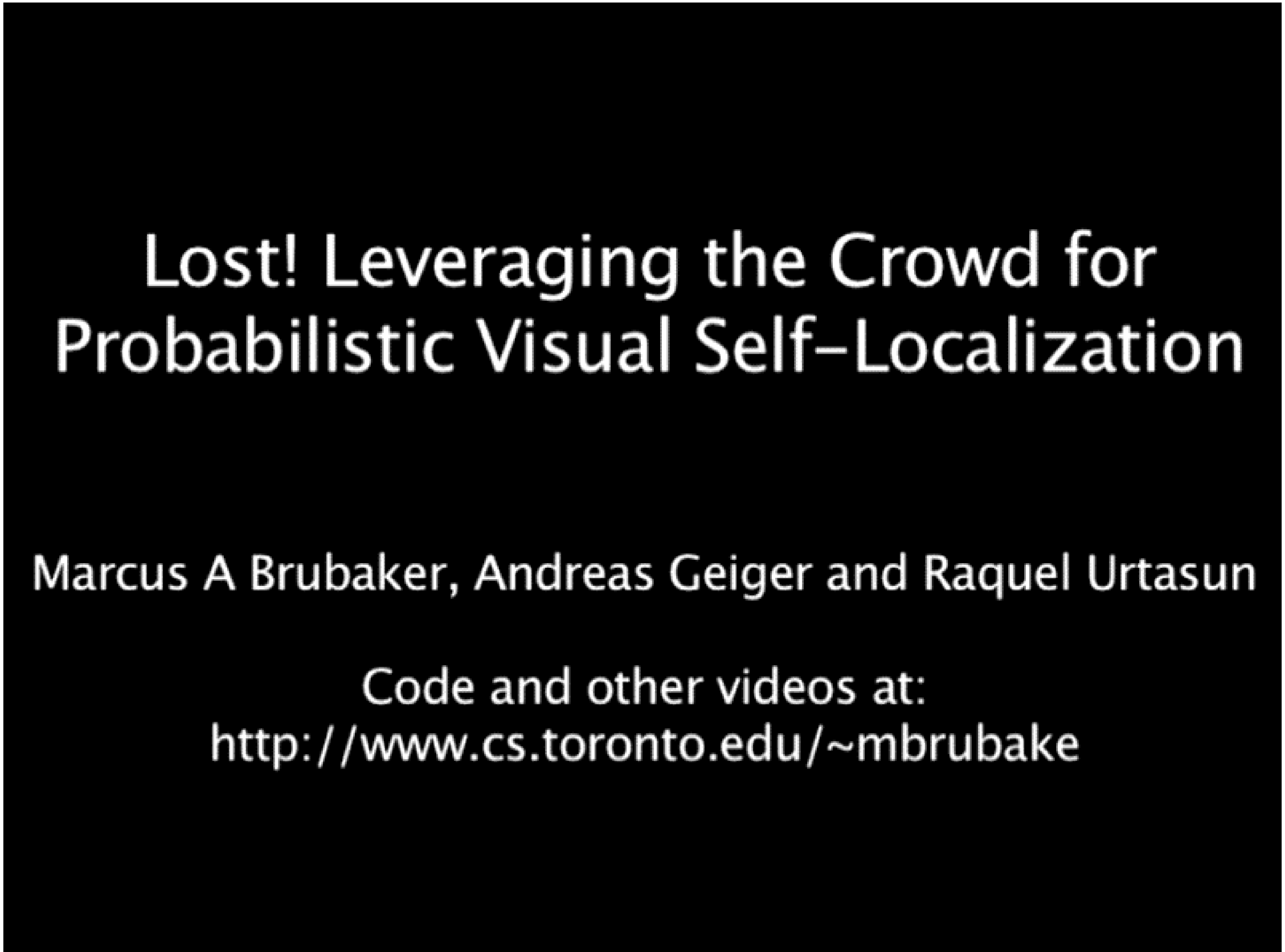
- Experiments: Ambiguous Sequences



In some circumstances,
geometric road shape + driving trajectory →
difficulties in localization

CVPR13.pdf (toronto.edu)
CVPR13-Slides.pdf (toronto.edu)
Lost! Leveraging the Crowd for Probabilistic Visual Self-Localization (toronto.edu)

Visual Odometry + Road Maps = Self-Localization



[CVPR13.pdf (toronto.edu)](CVPR13.pdf)

Summary: Vision-based Pose Estimation

- **Absolute** pose estimation: visual place recognition system
  - Image processing module
  - Mapping module
  - Belief generation module
- **Relative** motion estimation: Visual Odometry (VO)
  - Basic principles and pipeline for VO
  - Front-end and back-end of VO
  - Estimating the relative motion between two frames: feature-based methods and direct methods
- Examples and applications
  - Combining VO and road maps for self-localization

References

1. https://www.edmundoptics.com/knowledge-center/application-notes/imaging/understanding-camera-sensors-for-machine-vision-applications/
2. https://thinklucid.com/tech-briefs/understanding-digital-image-sensors/
3. cvp_the-fundamentals-of-camera-and-image-sensor-technology_jon-chouinard.pdf (visiononline.org)
4. Learn | Full Well Capacity and Pixel Saturation (princetoninstruments.com)
5. EMVA1288-3.0.pdf
6. Computer Vision on Rolling Shutter Cameras (liu.se)
7. F5e.fm (uio.no)
8. https://isl.stanford.edu/~abbas/ee392b/lect08.pdf
9. A Short Tutorial on Three-Dimensional Cameras (inrialpes.fr)
10. Introduction to Time-of-Flight Camera (Rev. B)
11. Systems (vision-doctor.com)
12. Introduction-to-Machine-Vision.pdf (assemblymag.com)
13. lecture-11-camera-isp.pdf (stanford.edu)
14. How to achieve 1-meter accuracy in Android | Coordinates.net
15. Inertial Sensors: IMU, GPS-INS, AHRS, MRU for UAVs & Robotics (unmannedsystemstechnology.com)