EECE5698
Networked XR Systems

Some slides credits: Anthony Rowe, Intro to XR course, CMU

Lecture Outline for Today

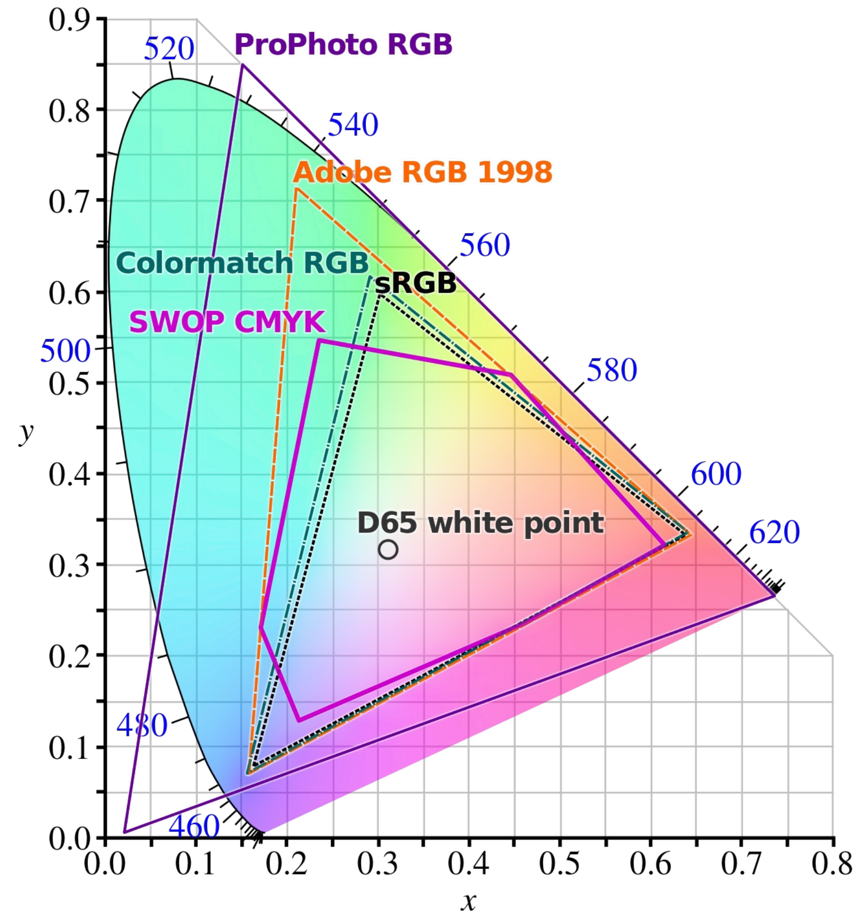
- Discuss Homework1
- XR Internals
- Sensors
- Sensing Algorithms

XR Internals

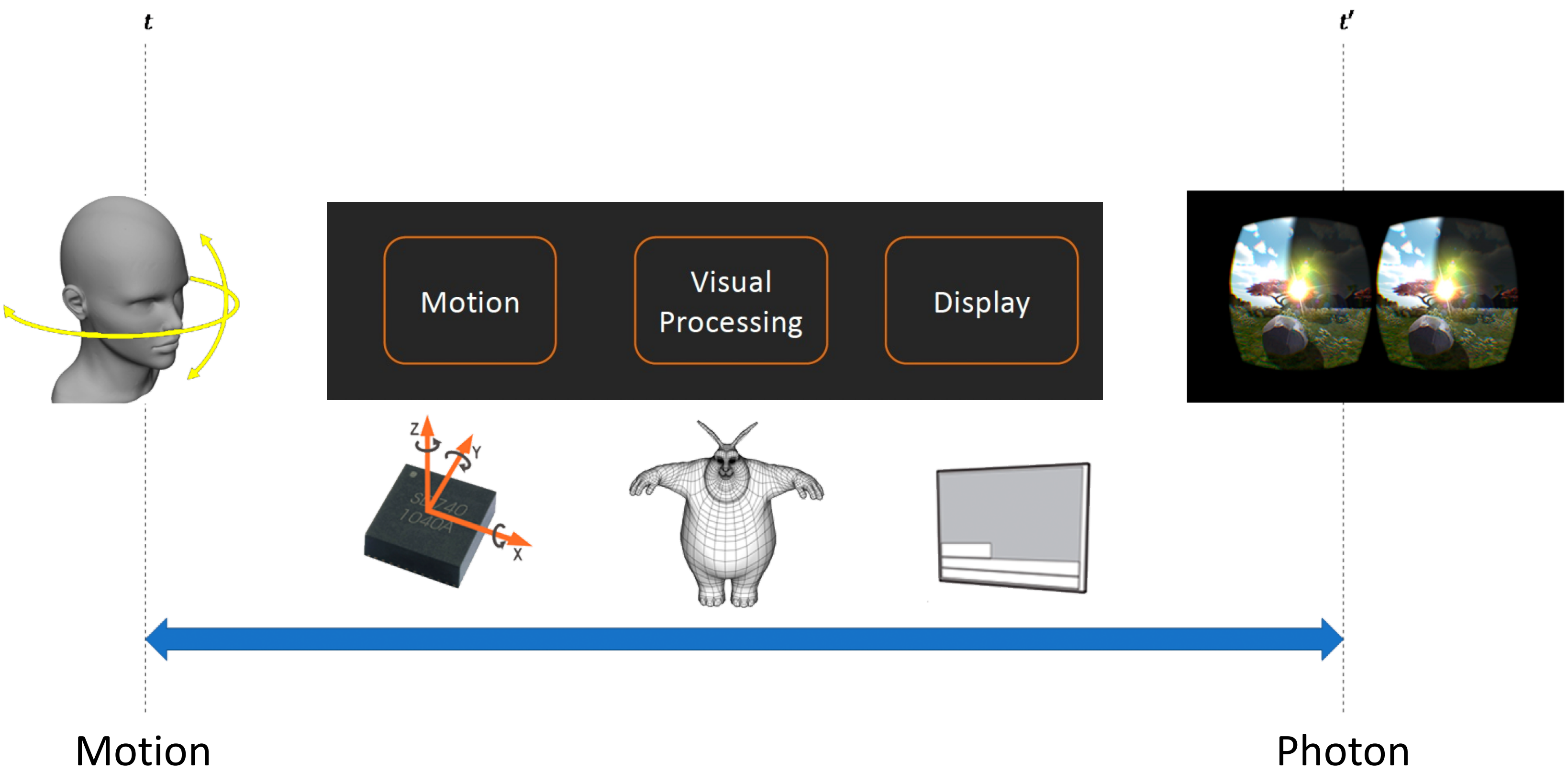
- Perception
- Motion to Photon Latency
- Real-time Rendering
- Sensors and Sensing Algorithms
    - Positioning and Tracking
    - 3D Reconstruction

XR Perception

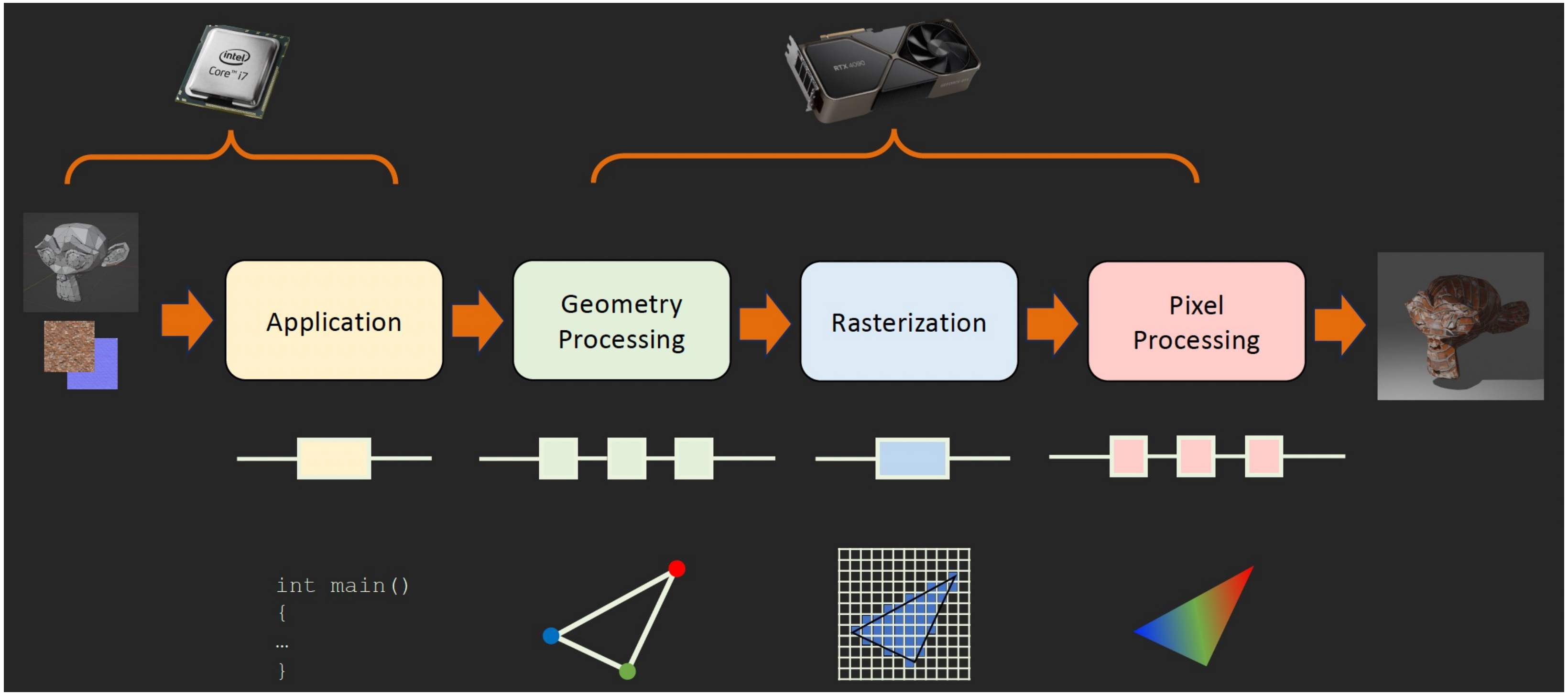
- Visual
  - Color
  - Quality/spatial resolution
  - Depth resolution
  - Temporal resolution
  - Field of view
- Non-visual
  - Sense of touch
  - Audio
  - Balance
  - Smell



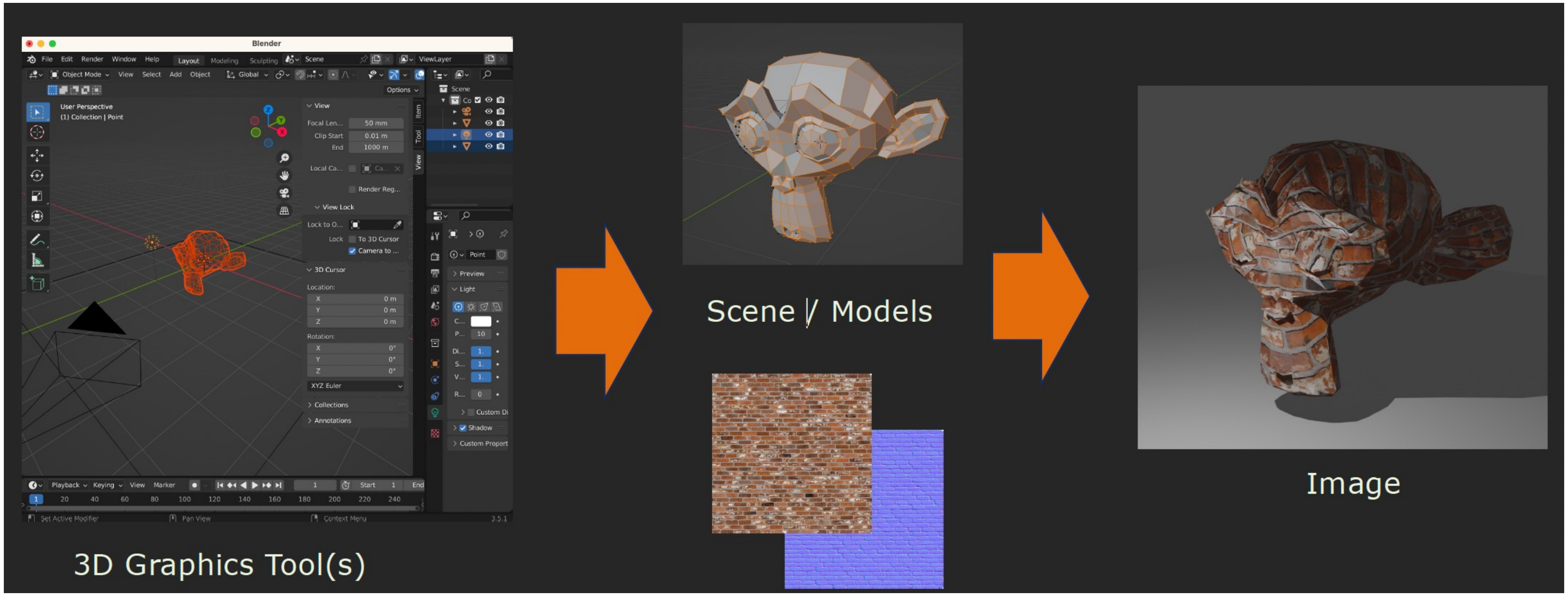
XR Perception: Motion to Photon Latency



Real-time Rendering



Real-time Rendering



Sensors and Sensing Algorithms

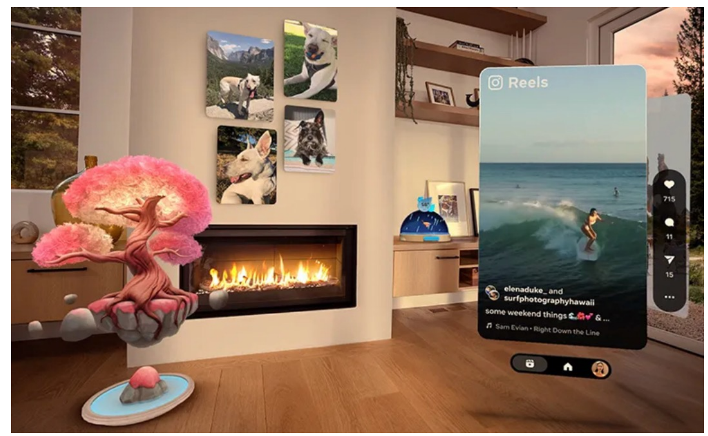
- Popular Sensors
  - Color camera
  - Depth camera
  - Microphone
  - Inertial
  - Gyro
  - RF
- Functionality
  - Positioning and Tracking
  - 3D Scene Reconstruction

Positioning and Tracking

- What to position and track?
  - Users
    - Hands
    - Face
    - Eyes
    - Head
    - Body
    - Activity
    - Physiological signals
  - Environment
    - Objects

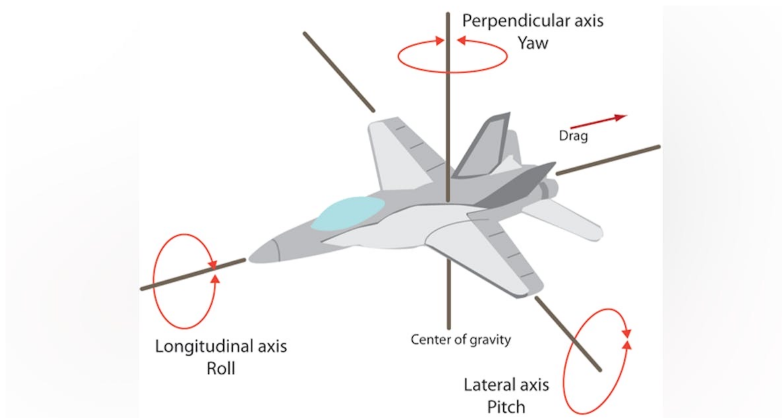
Positioning and Tracking

- Why do we need it?
  - For view port control
  - Place virtual content
  - Interact with virtual content
  - Occlusion
  - Adaptive rendering
  - Persistent anchors
  - And more…

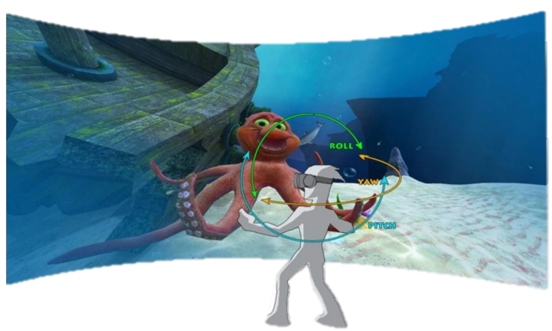


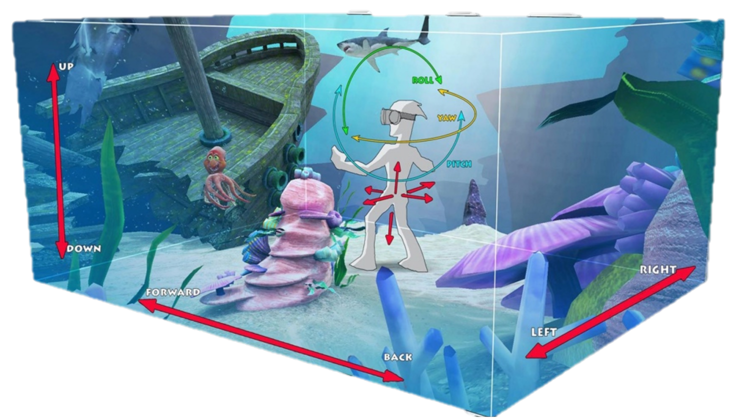
Positioning and Tracking

- You need to know where you are in the world
  - GPS?
  - Visual
  - Inertial
  - Lidar
  - RF



- 3-DoF
- 6-DoF

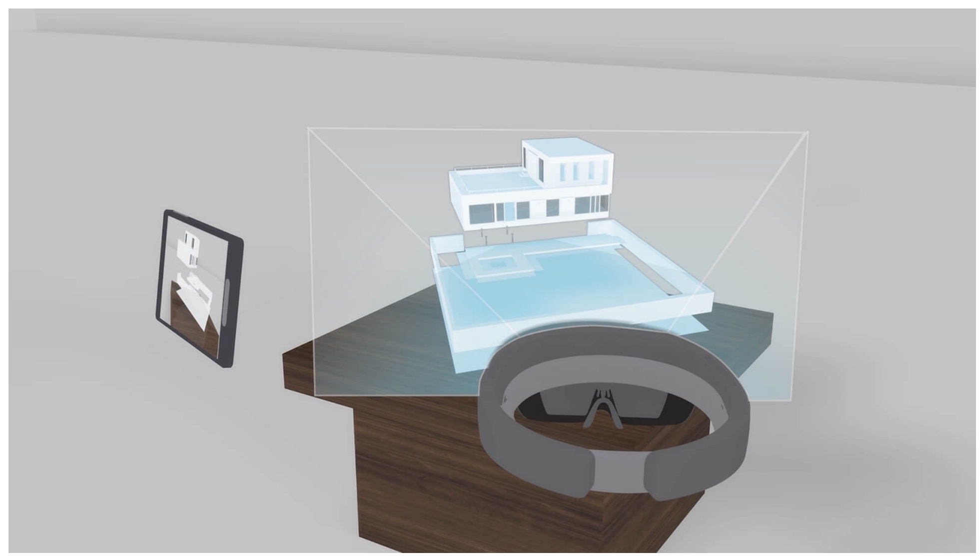




X, Y, Z & Yaw, Pitch, Roll

Positioning and Tracking

- Anchors
  - Anchors ensure that objects appear to stay at the same position and orientation in space, helping you maintain the illusion of virtual objects placed in the real world.



  - Plane
  - Wall
  - Floor
  - Face…
  - Anything that you can identify well

Positioning and Tracking

- Typical metrics of importance
  - Accuracy
  - Latency
  - Tracking drift
  - Tracking jitter
  - Update rate
  - Reliability

Visual Positioning & Tracking

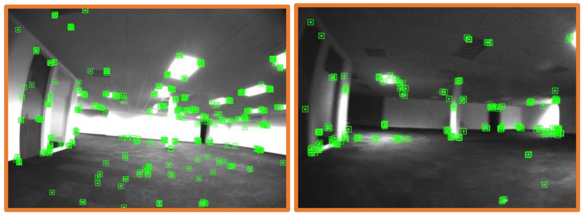
- Step1: Capture images
  - Mono or Stereo or multiple cameras
- Step2: Feature Extraction
  - Features are detected in the first frame, and then matched in the second frame.

(a) Well-lit (568 matches)

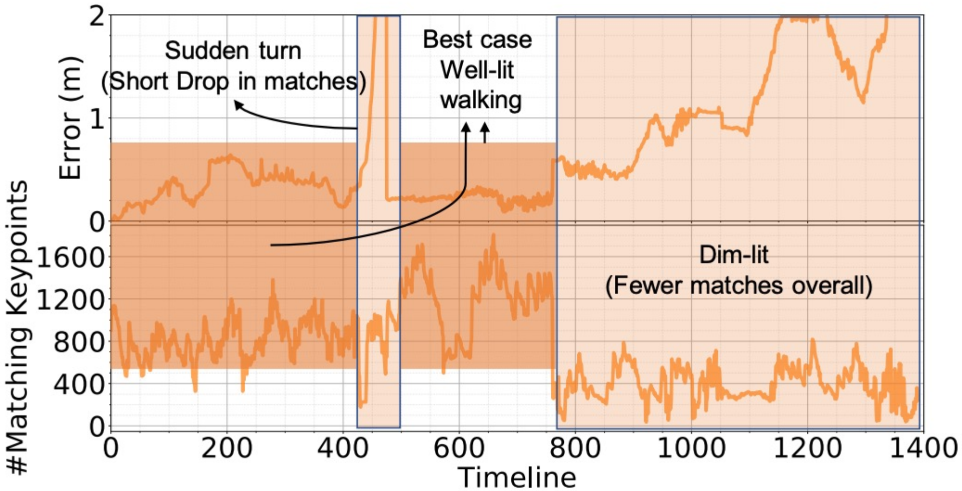
(b) Dim-lit (252 matches)

Visual Positioning & Tracking

- Step1: Capture images
  - Mono or Stereo or multiple cameras
- Step2: Feature Extraction
  - Features are detected in the first frame, and then matched in the second frame



(a) Well-lit (568 matches) (b) Dim-lit (252 matches)



(c) Matches vs. Error

Visual Positioning & Tracking

- Step3: Optical flow estimation

Get rid of outliers

Visual Positioning & Tracking

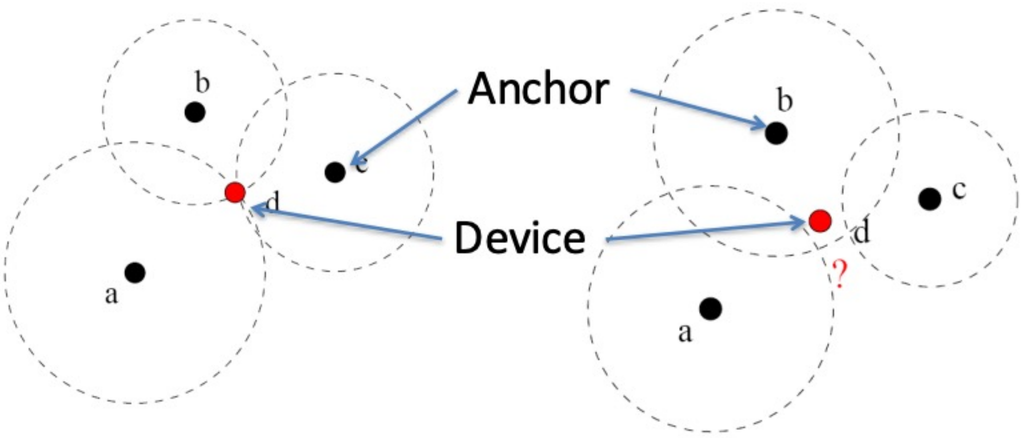
- Step4: Estimate camera motion from optical flow
  - The optical flow field illustrates how features diverge from a single point, the *focus of expansion*. The focus of expansion can be detected from the optical flow field, indicating the direction of the motion of the camera, and thus providing an estimate of the camera motion.

Visual Positioning & Tracking

- Limitations:
  - Heavily depends on the environment
    - Lighting conditions
    - Geometry of the objects in the environment
    - Uniform surfaces or color
    - Moving objects
    - Fails when too close to objects; camera view occluded

RF-based Tracking

- Range based tracking
  - Convert received signal strength (RSS) or signal timing to a distance estimate with respect to anchor nodes with known locations.
  - Problem: distance estimates may be erroneous, and the circles may not intersect at a single point.



RF-based Tracking

How to estimate location when the circles do not intersect?

Idea: localize at a point that presents the minimum error to the circles by some reasonable error measure.

k anchors at positions (x_i, y_i)

Assume node to be localized has actual location at (x_0, y_0)

Distance estimate between node 0 and anchor *i* is r_i

Error:

$$f_i = r_i - \sqrt{(x_i - x_0)^2 + (y_i - y_0)^2}$$

RF-based Tracking

Linearization and Min Mean Square Estimate

- Ideally, we would like the error to be 0

$$f_i = r_i - \sqrt{(x_i - x_0)^2 + (y_i - y_0)^2} = 0$$

- Re-arrange:

$$(x_0^{\ 2} + y_0^{\ 2}) + x_0(-2x_i) + y_0(-2y_i) - r_i^{\ 2} = -x_i^2 - y_i^2$$

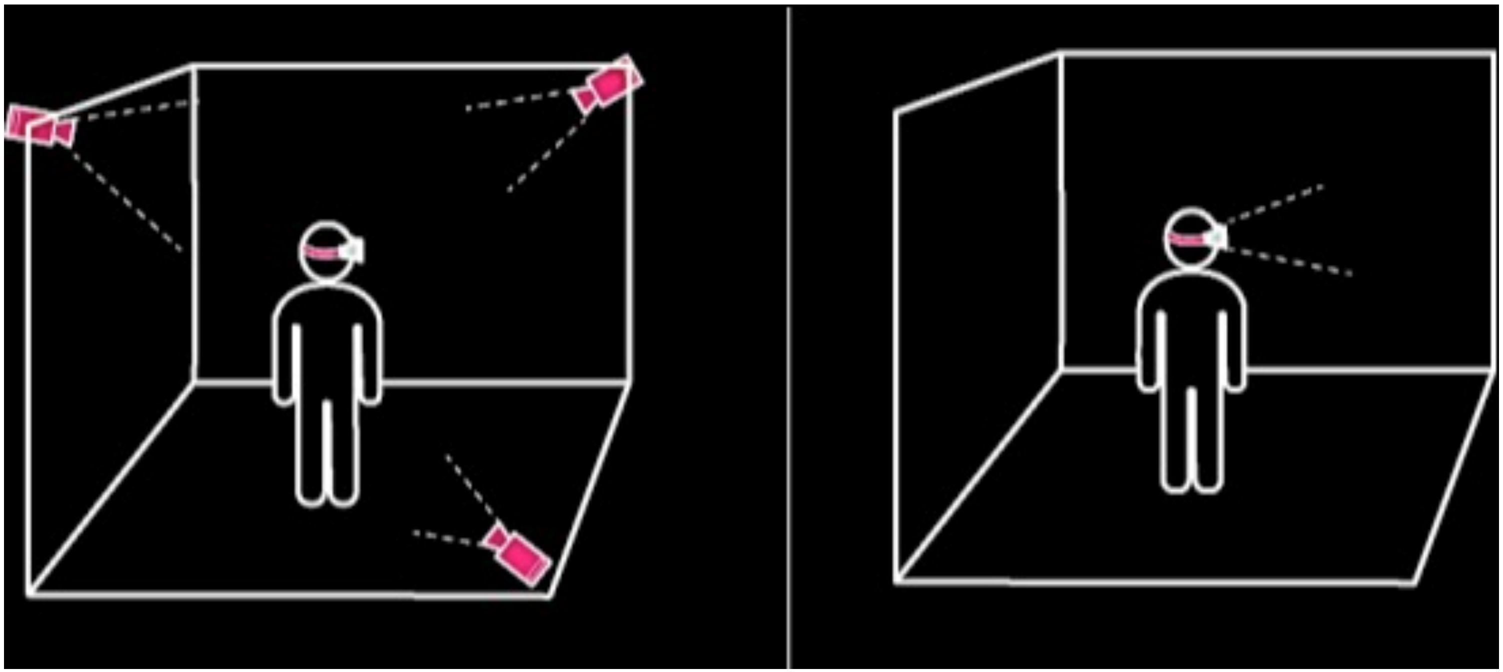
- Subtract the last equation from the previous ones to get rid of quadratic terms.

$$2x_0(x_k - x_i) + 2y_0(y_k - y_i) = r_i^2 - r_k^{\ 2} - x_i^2 - y_i^2 + x_k^2 + y_k^2$$

- Note that this is linear.

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

Outside in and Inside out Tracking



Outside in Inside out

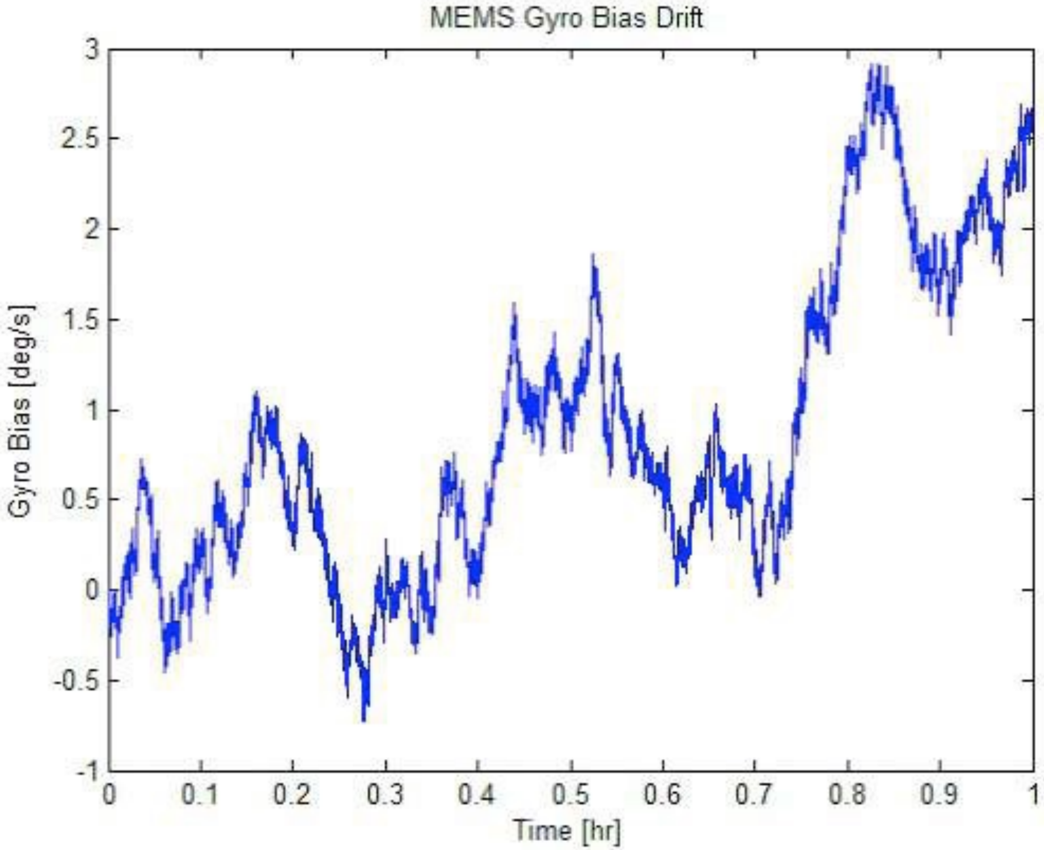
Inertial sensing

- Accelerometer & Gyroscope
  - Measuring linear acceleration (accelerometer) and / or angular orientation rates (gyroscope)
  - No transmitter, cheap, small, high frequency, wireless

https://youtu.be/-0hSQFbt67U?t=24

Inertial sensing

- Accelerometer & Gyroscope
  - Drift

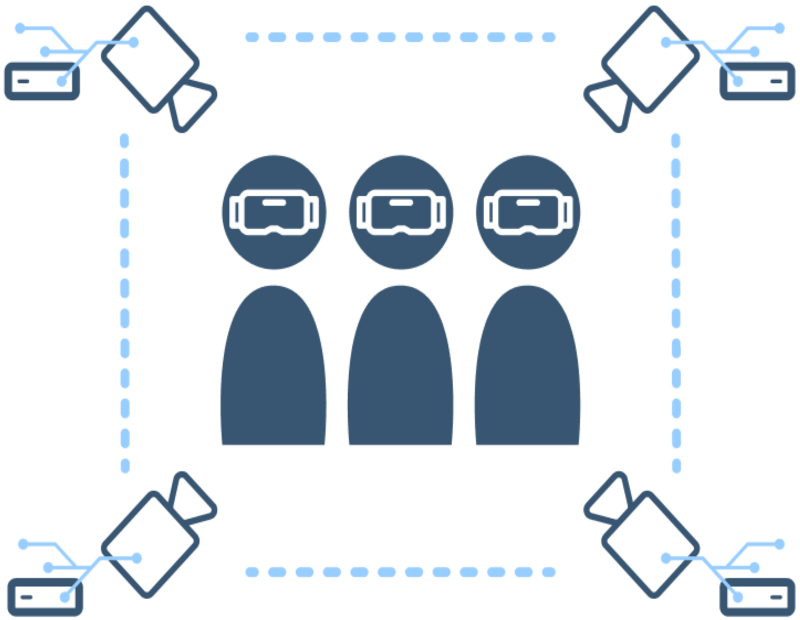


3D Reconstruction

- Camera Calibration
- Depth Sensing
- Surface Extraction
- Texture Generation

3D Reconstruction

- Camera Calibration
  - Multiple cameras
  - Distortion
  - Intrinsic and extrinsic parameters are different for different cameras



3D Reconstruction

- Camera Calibration

- **Input**: set of pictures
- **Output**: camera position, orientation, intrinsic parameters (focal length, optical center)

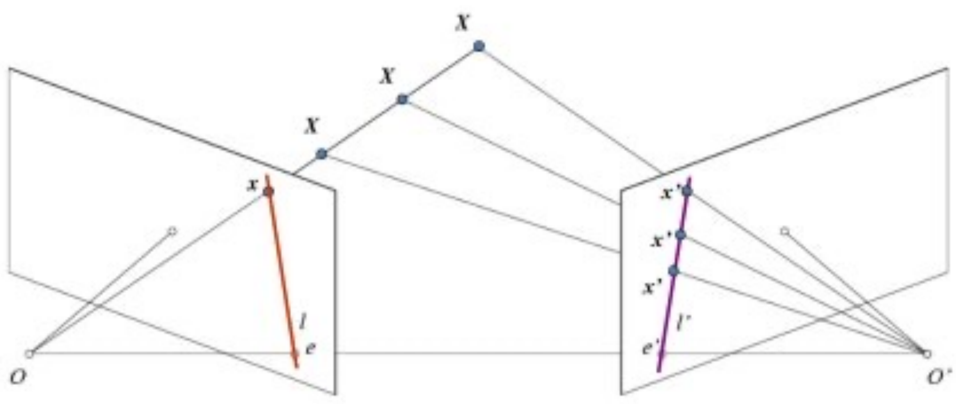
3D Reconstruction

- Depth Sensing
  - Input: set of calibrated images
  - Output: distance to object for each pixel in the image

- Popular methods
  - Stereo triangulation
  - Time of flight
  - Structured light projection

3D Reconstruction

- Depth Sensing

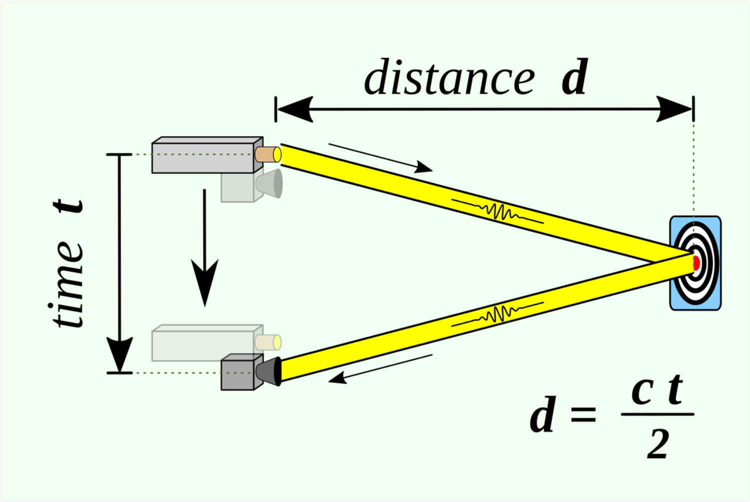


Stereo Triangulation

Zed Camera

3D Reconstruction

- Depth Sensing

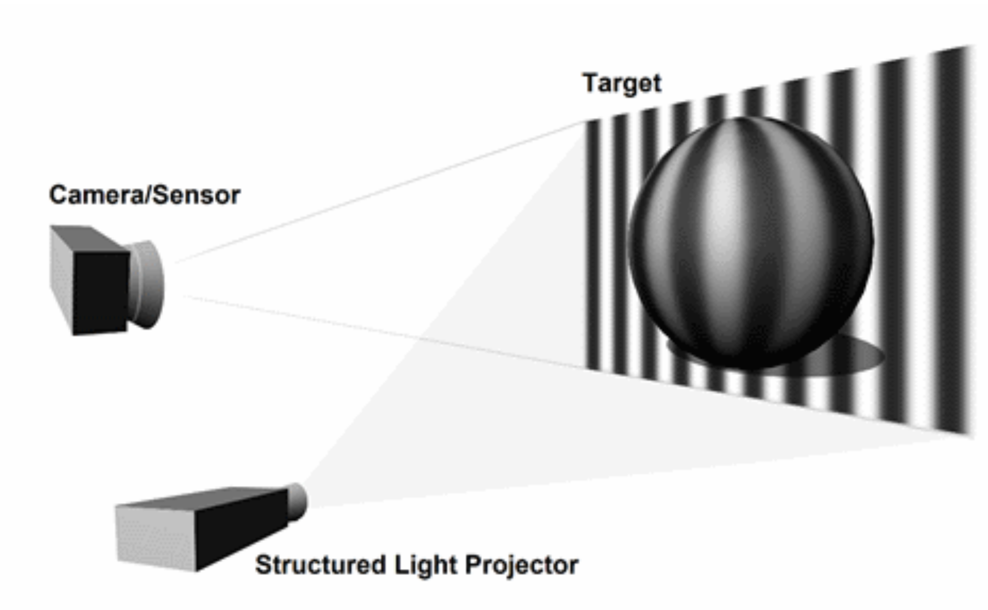


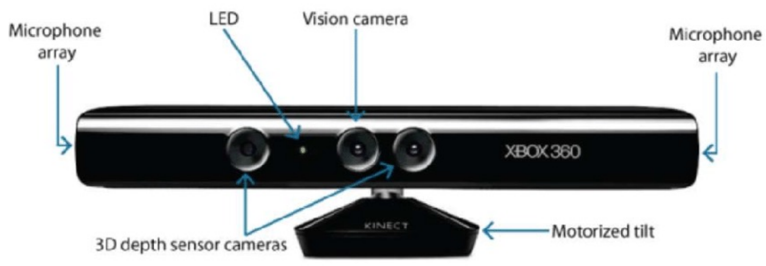
Time of flight

Helios

3D Reconstruction

- Depth Sensing





Azure Kinectv1

Structured light projection

3D Reconstruction

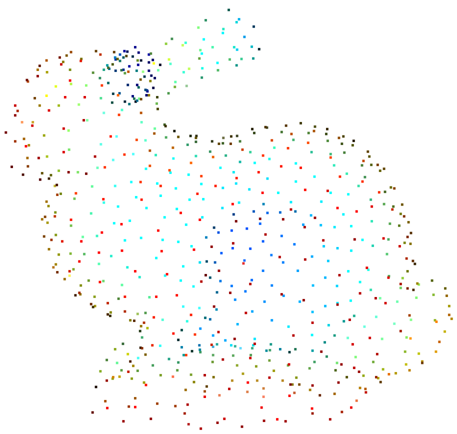
- Depth Sensing

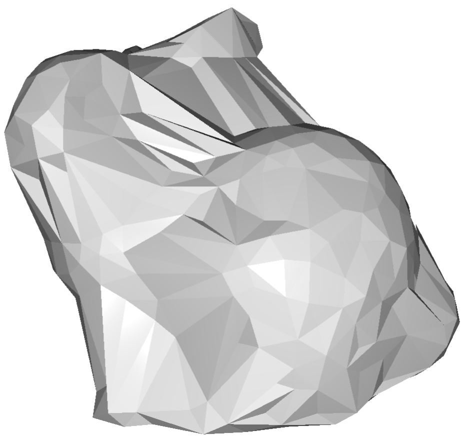
	Stereo vision	Structured light	Laser triangulation	Time of Flight
Distance & range	Medium to far (depending on the distance of the 2 cameras) & limited 2m to 5m	Short to medium & scalable cm to 2m	Short & Limited cms	Far & scalable 30-50cm to 20-50m
Resolution	Medium	Medium	Varies	High
Depth accuracy	Medium	Medium to very high in short range	Very high	Medium
Software complexity	High	Medium	High	Low
Real-time capability	Low	Low	Low	High
Low light behaviour	Weak	Good	Good	Good
Outdoor light	Good	Weak	Weak	Weak to good
Compactness	Medium	Medium	Medium	Very compact
Material costs	Low	High	High	Medium
Total operating cost (including calibration efforts)	High	Medium to high	High	Medium

https://www.azom.com/article.aspx?ArticleID=16003

3D Reconstruction

- Surface Extraction from Depth
  - Input: set of calibrated images & depth maps
  - Output: mesh of object





3D Reconstruction

- Texture Generation
  - Input: set of calibrated images and mesh of object
  - Output: atlas and texture

Summary of the Lecture

- Discuss Homework1
- XR Internals
- Sensors
- Sensing Algorithms

Next Up: XR Data Structures