Intelligent systems
An introduction to 3D computer vision
(From 3D Reconstruction to Recognition)

What are intelligent systems?

Systems that can analyse digital information and sensory data (KNOWLEDGE), draw the right conclusions (REASONING), and act on them in the digital or physical world[1].

Intelligent learning and analysis systems

Systems able to analyse large amounts of data (e.g., Expert systems, Machine learning-based inferences).

Intelligent sensorimotor (mobile) systems

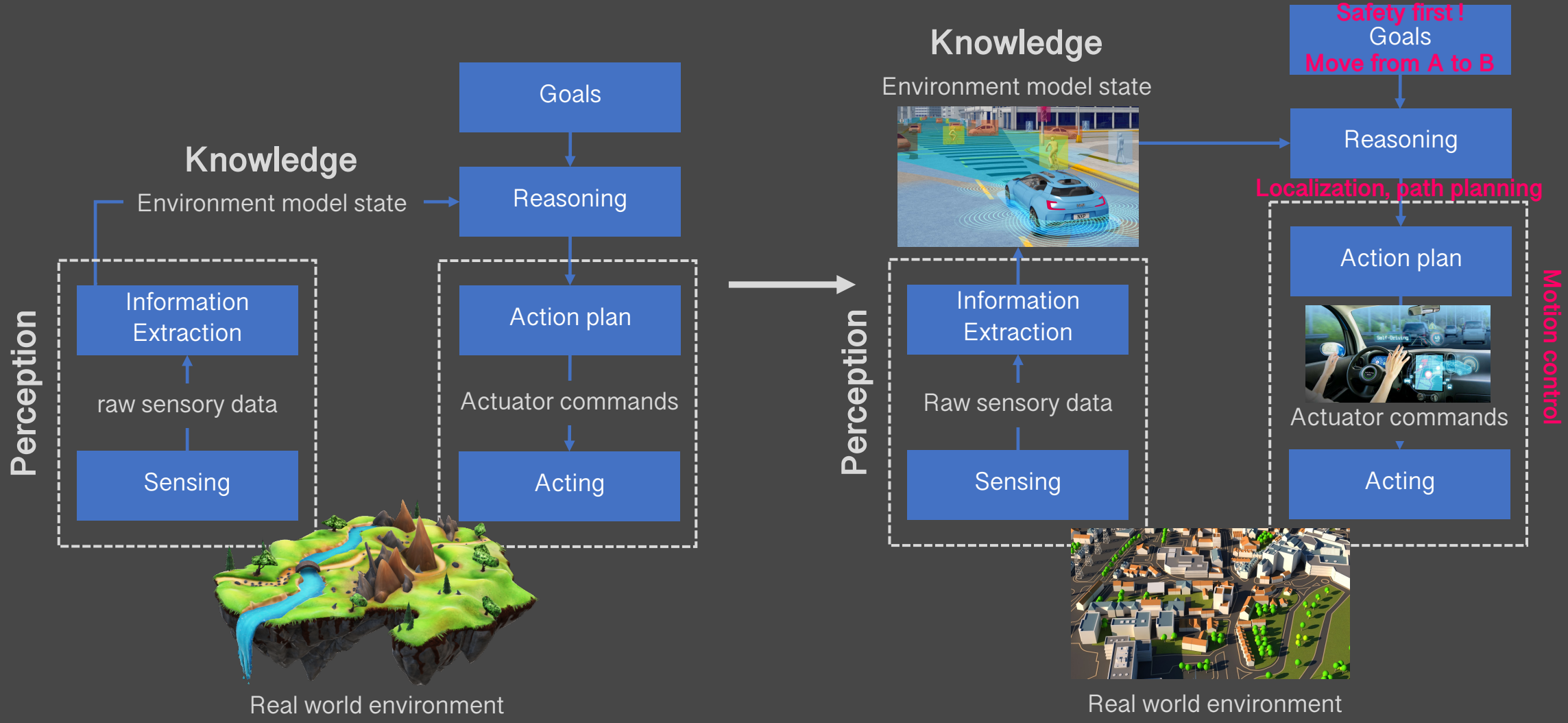
Systems that perceive the physical world and act on it in order to achieve some goals.

Autonomous systems

Intelligent systems that act without recourse to human control.

[1] https://www.informatik.uni-bonn.de/en/research-and-phd/intelligent-systems

Intelligent mobile systems



Challenges in intelligent mobile systems

Uncertainty

Physical sensors provide limited, noisy and inaccurate information. Consequently, any action taken may be incorrect !

Dimension reduction

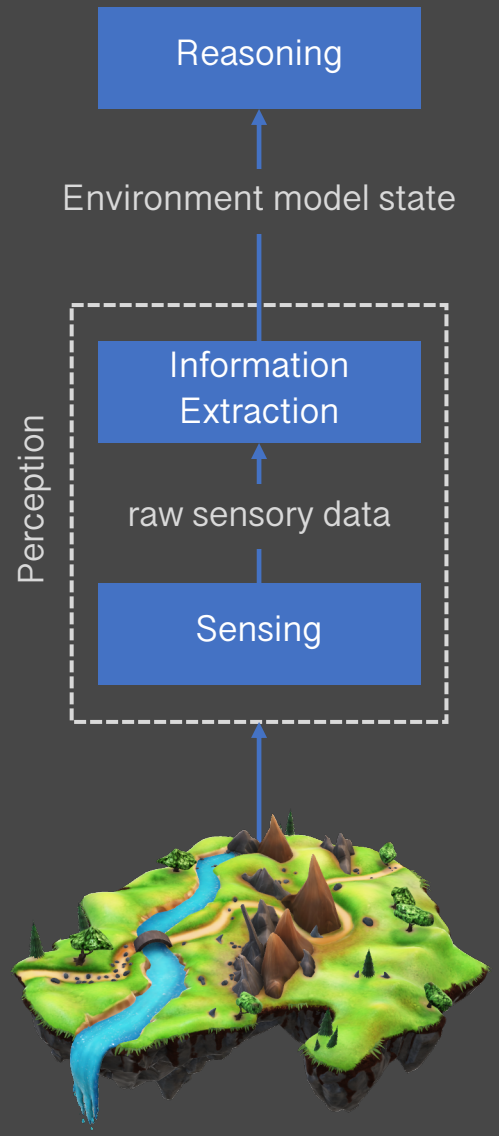
The physical environment is multi-dimensional whereas sensors are not !

Time-consuming computations

Information extraction and reasoning require extensive computations.

Dynamic world

Surrounding environment perception and decision tasks have to be achieved faster than the changes in the environment …



Vision processing in intelligent mobile systems

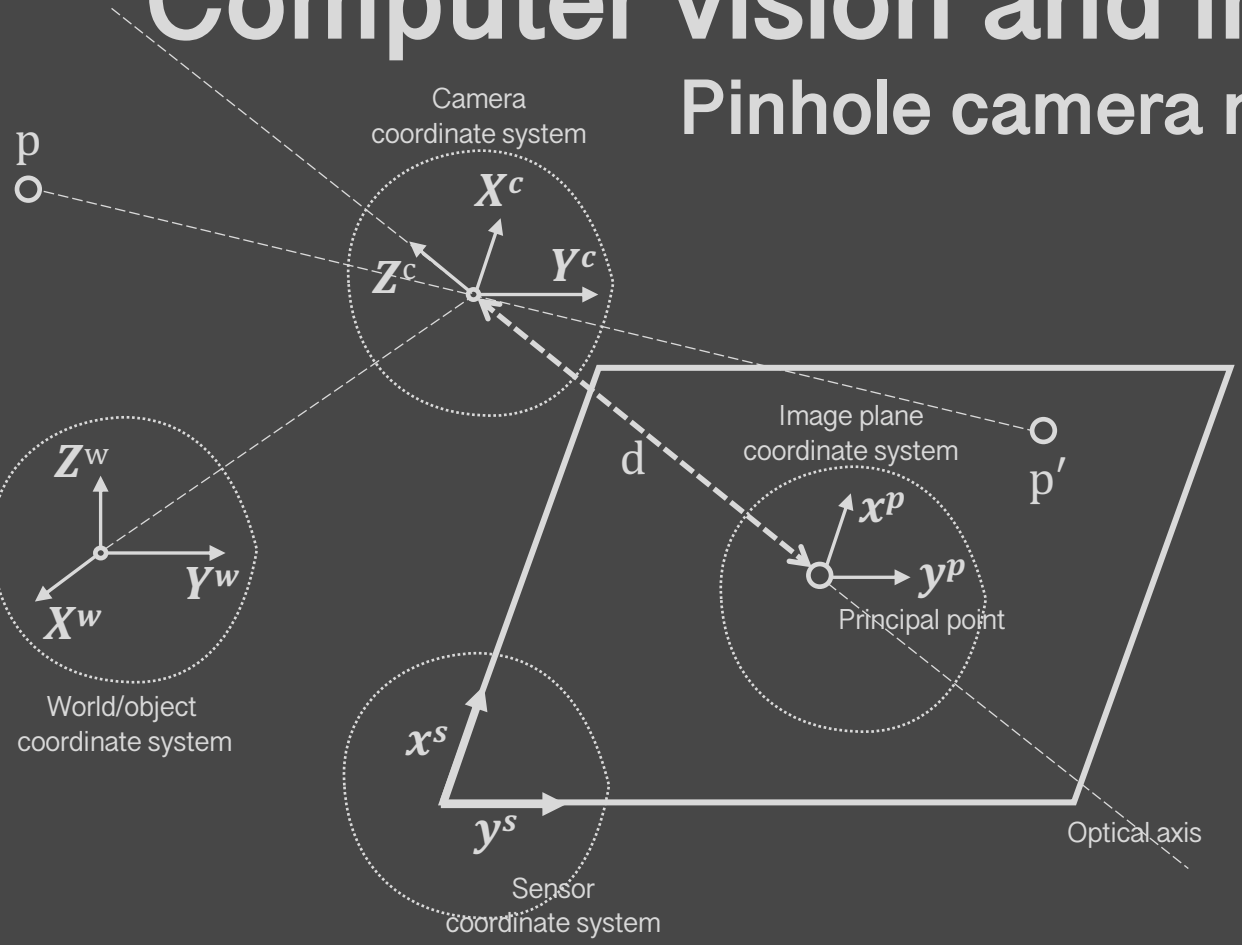
Vision processing enables intelligent systems to identify, navigate, inspect or handle parts or tasks (e.g. product sorting, product assembly, measuring, etc.).

3D vs 2D vision processing

- Thickness, height and volume **measurement**,
- Product quality **inspection**,
- **Guidance** and surface tracking,
- Bin **picking** for placing, packing or assembly,
- Environment **scanning** and **digitization**,
- **Obstacle avoidance**,
- …

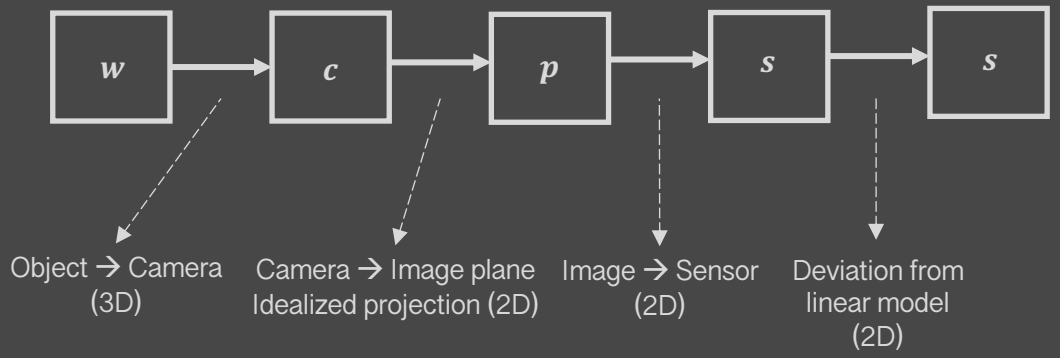
Computer vision and image processing

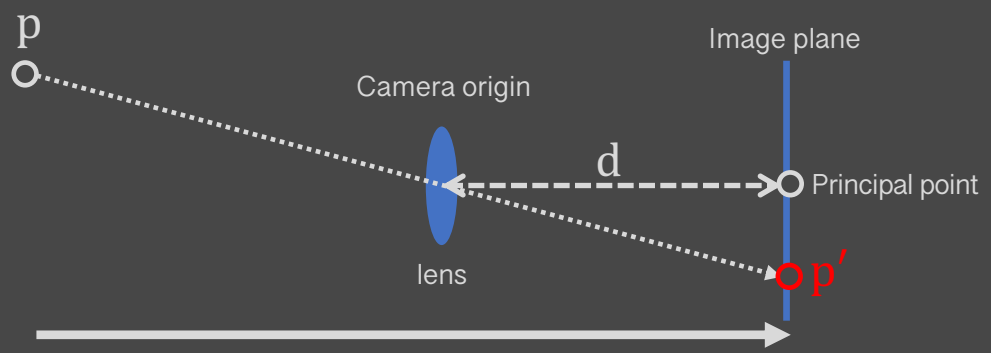
Pinhole camera model (perspective camera model)



Pixel coordinates $\begin{bmatrix} x^s \\ y^s \end{bmatrix} = T \begin{bmatrix} X^w \\ Y^w \\ Z^w \end{bmatrix}$ World coordinates

$$T \cong K[R|t]$$





Sources

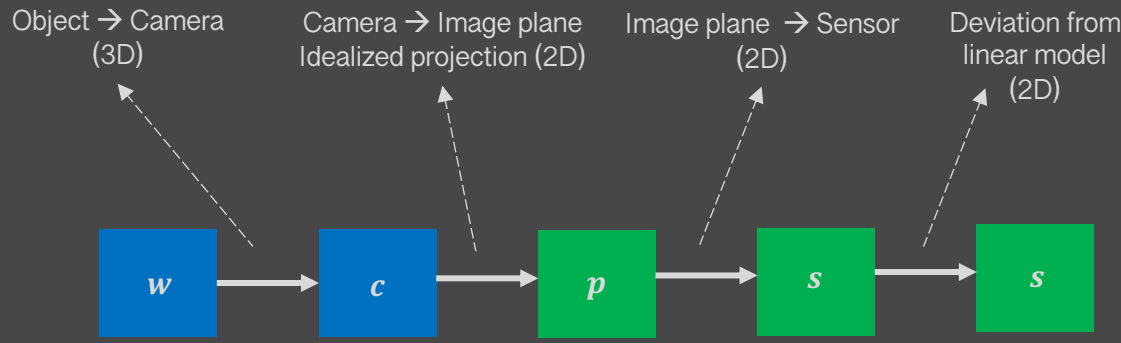
https://cvgl.stanford.edu/teaching/cs231a_winter1415/lecture/lecture2_camera_models_note.pdf

Cyrill Stachniss, cyrill.stachniss@igg.uni-bonn.de

https://www.ipb.uni-bonn.de/html/teaching/photo12-2021/2021-pho1-20-camera-params.pptx.pdf

Computer vision and image processing

Pinhole camera model (perspective camera model)



$$T \cong K[R|T]$$

Extrinsics Intrinsics

Extrinsic parameters

Describe the pose of the camera with respect to the world.

Intrinsic parameters

The process of projecting points from the 3d camera coordinates to the 2d sensor image coordinates.

Sources

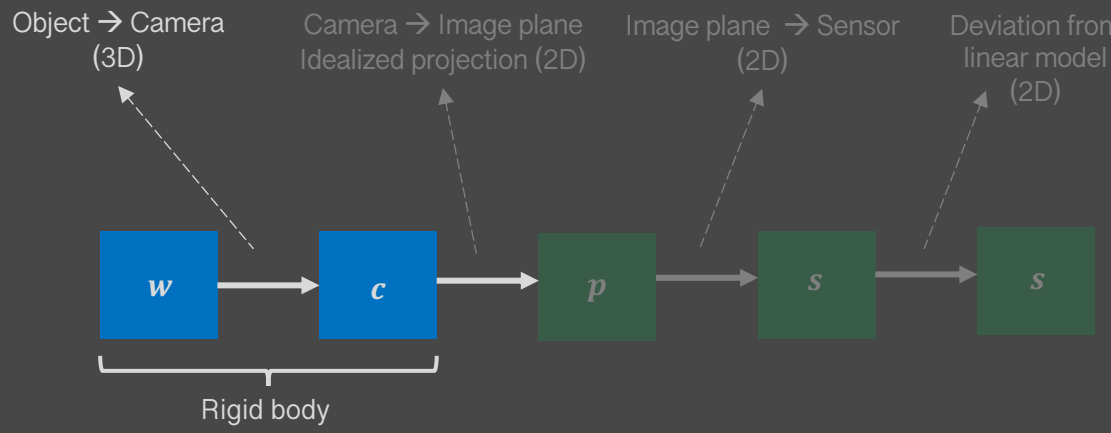
https://cvgl.stanford.edu/teaching/cs231a_winter1415/lecture/lecture2_camera_models_note.pdf

Cyrill Stachniss, cyrill.stachniss@igg.uni-bonn.de

https://www.ipb.uni-bonn.de/html/teaching/photo12-2021/2021-pho1-20-camera-params.pptx.pdf

Computer vision and image processing

Pinhole camera model (perspective camera model)

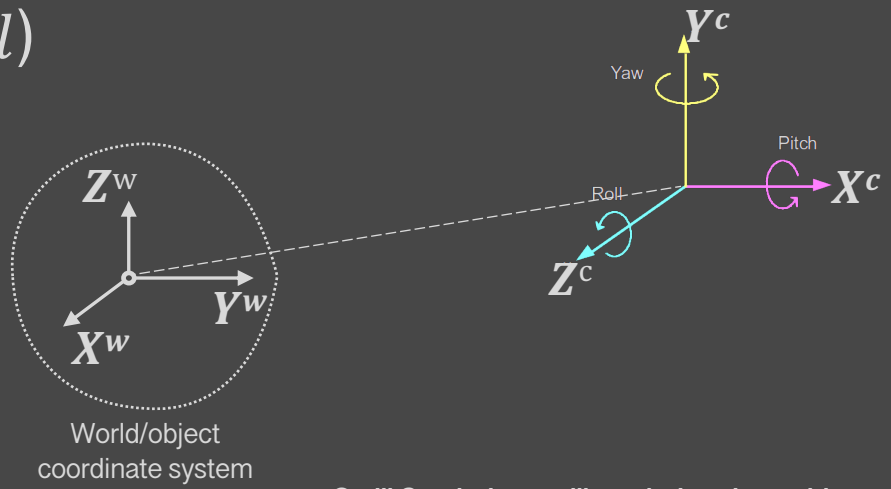


Extrinsic parameters

Extrinsics

Intrinsics

6 parameters : 3 for the position (x^c, yc, zc) relative to the world coordinates origin, 3 for the heading $(yaw, pitch, roll)$



Sources

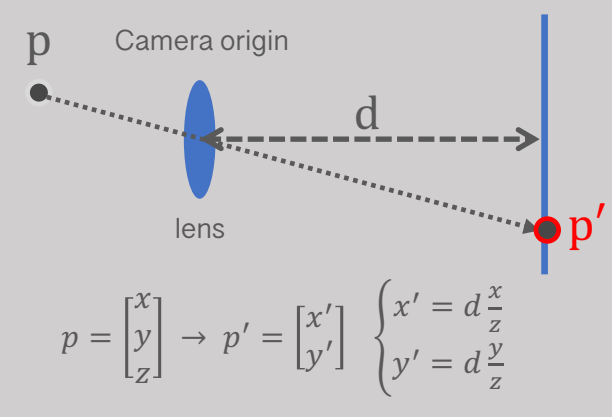
https://cvgl.stanford.edu/teaching/cs231a_winter1415/lecture/lecture2_camera_models_note.pdf

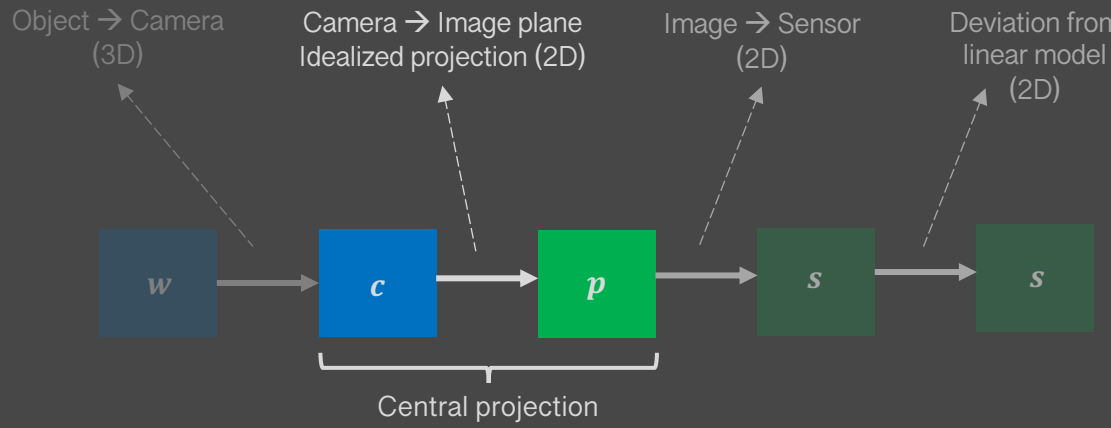
Cyrill Stachniss, cyrill.stachniss@igg.uni-bonn.de

https://www.ipb.uni-bonn.de/html/teaching/photo12-2021/2021-pho1-20-camera-params.pptx.pdf

Computer vision and image processing

Pinhole camera model (perspective camera model)





Intrinsic parameters

1. ***Ideal perspective projection*** to the image plane
   a) Distortion-free lens,
   b) All rays are straight lines and pass through the projection center (origin of the camera coordinate system),
   c) Focal point and principal point lie on the optical axis,
   d) The distance from the camera origin to the image plane is the constant d (focal length)

Sources

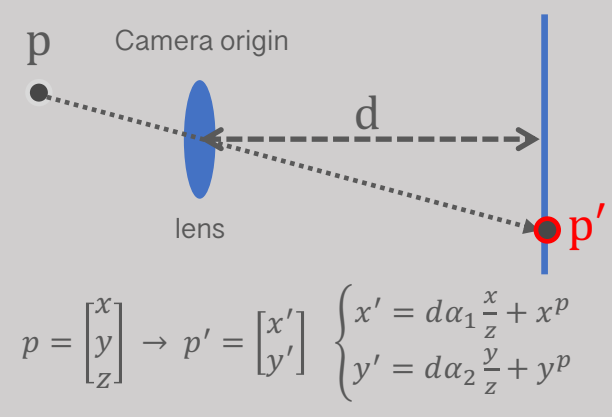
https://cvgl.stanford.edu/teaching/cs231a_winter1415/lecture/lecture2_camera_models_note.pdf

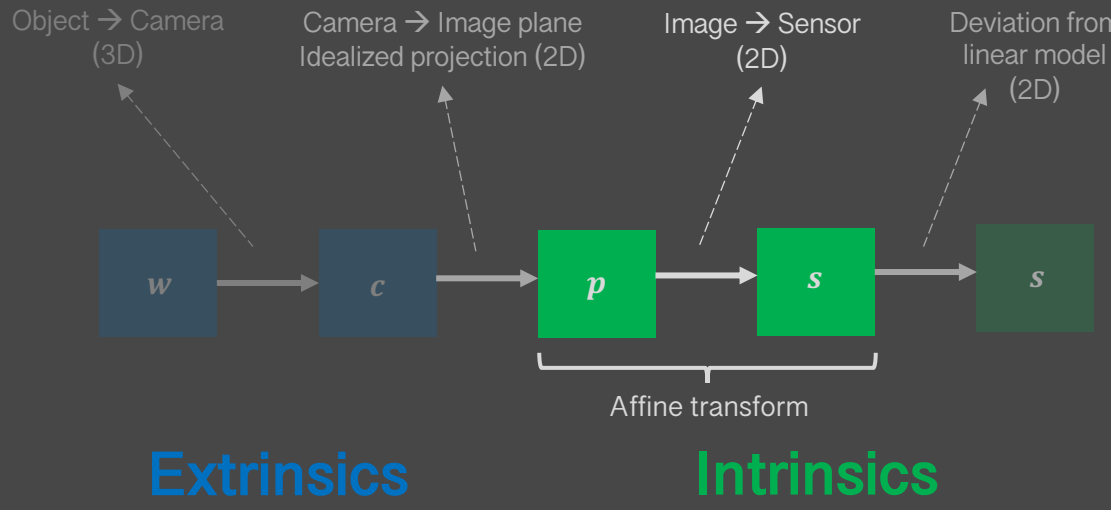
Cyrill Stachniss, cyrill.stachniss@igg.uni-bonn.de

https://www.ipb.uni-bonn.de/html/teaching/photo12-2021/2021-pho1-20-camera-params.pptx.pdf

Computer vision and image processing

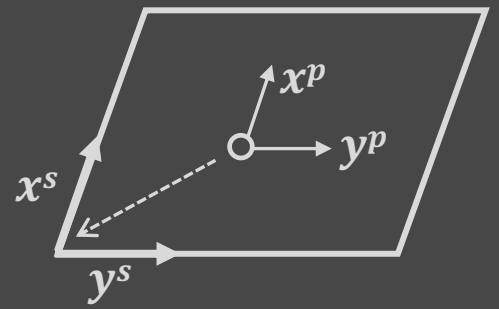
Pinhole camera model (perspective camera model)





Intrinsic parameters

2. ***Mapping from the image plane to the sensor (linear errors)***



   a) The origin of the sensor system is not at the principal point → Compensation through a translation,
   b) Apply aspect ratio α_1 and α_2 (=1 unless pixels are not square),
   c) Axis skew causes shear distortion in the projected image that needs to be compensated by s (≈ 0)

Sources

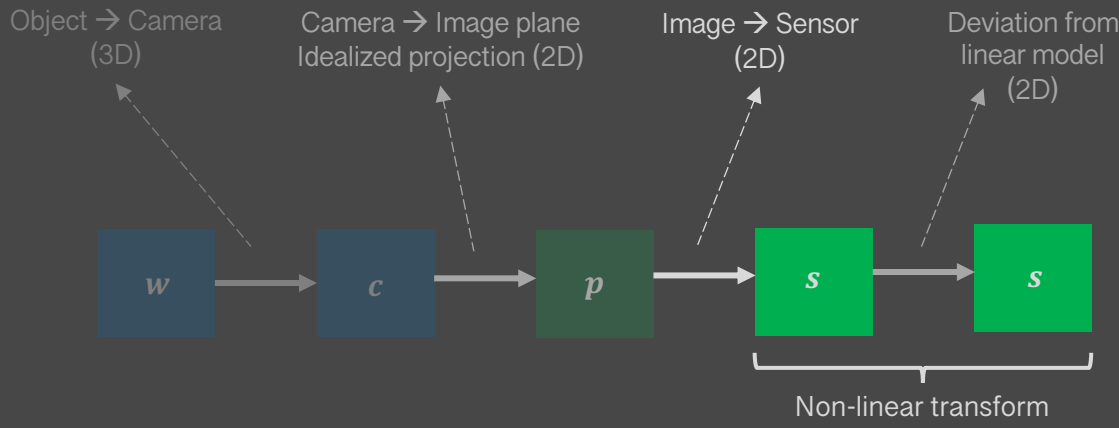
https://cvgl.stanford.edu/teaching/cs231a_winter1415/lecture/lecture2_camera_models_note.pdf

Cyrill Stachniss, cyrill.stachniss@igg.uni-bonn.de

https://www.ipb.uni-bonn.de/html/teaching/photo12-2021/2021-pho1-20-camera-params.pptx.pdf

Computer vision and image processing

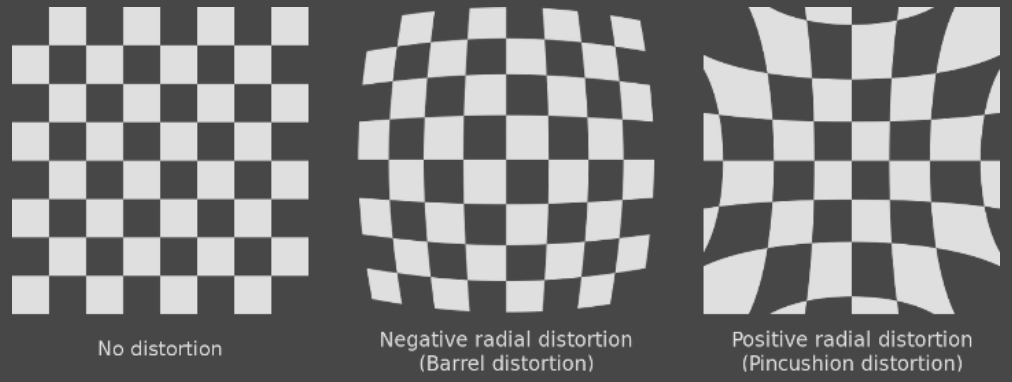
Pinhole camera model (perspective camera model)



Intrinsic parameters

3. *Handling non linearities*
   a) Lens imperfections,
   b) Planarity of the sensor,
   c) etc.

   ➔ Correction parameters



Sources

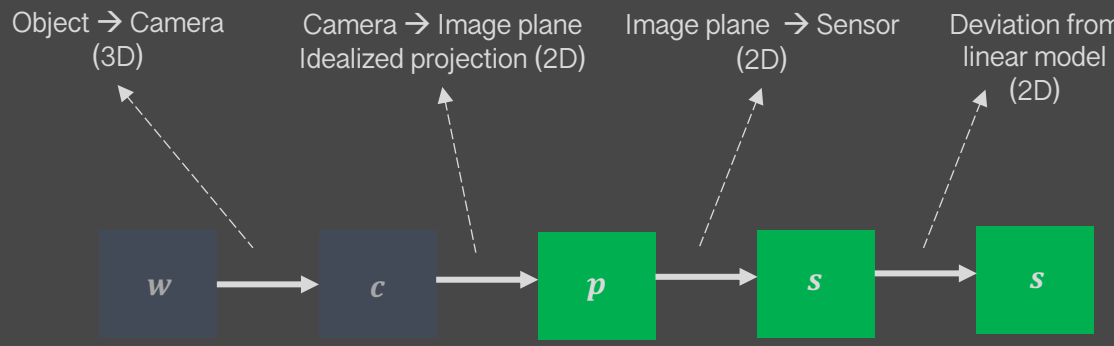
https://cvgl.stanford.edu/teaching/cs231a_winter1415/lecture/lecture2_camera_models_note.pdf

Cyrill Stachniss, cyrill.stachniss@igg.uni-bonn.de

https://www.ipb.uni-bonn.de/html/teaching/photo12-2021/2021-pho1-20-camera-params.pptx.pdf

Computer vision and image processing

Pinhole camera model (perspective camera model)



Extrinsics Intrinsics

Calibration matrix

If the intrinsics are **known**, the camera is said **calibrated with, considering homogeneous coordinate system :**

$$\begin{bmatrix} x^s \\ y^s \\ Z^w \end{bmatrix} = \overbrace{\begin{bmatrix} \alpha_1 d & s & x^p \\ 0 & \alpha_2 d & y^p \\ 0 & 0 & 1 \end{bmatrix}}^{K} \overbrace{\begin{bmatrix} I & 0 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & x^c \\ r_{21} & r_{22} & r_{23} & y^c \\ r_{31} & r_{32} & r_{33} & z^c \\ 0 & 0 & 0 & 1 \end{bmatrix}}^{[R \mid t]} \begin{bmatrix} X^w \\ Y^w \\ Z^w \\ 1 \end{bmatrix}$$

Sources

https://cvgl.stanford.edu/teaching/cs231a_winter1415/lecture/lecture2_camera_models_note.pdf

Cyrill Stachniss, cyrill.stachniss@igg.uni-bonn.de

https://www.ipb.uni-bonn.de/html/teaching/photo12-2021/2021-pho1-20-camera-params.pptx.pdf

Computer vision and image processing

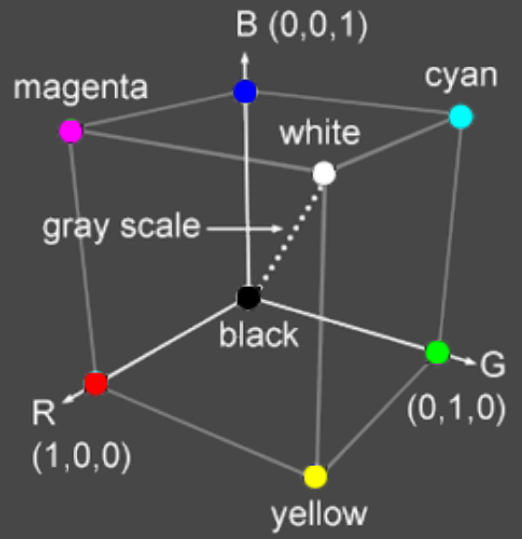
Color camera imply a color space approximation, from high spectral color space to lower spectral color sub-spaces.

Grayscale color model

One byte per pixel (values from 0 to 255) representing intensity of whites.

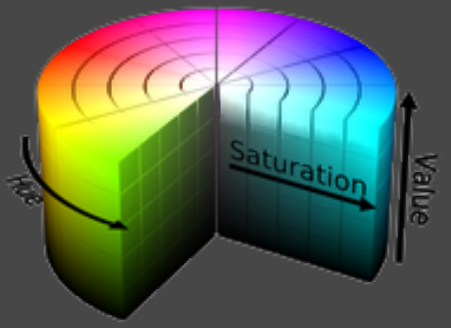
RGB color model

Three bytes per pixel, each representing intensity of red, green and blue (16581375 distinct colors). Only captures a small part of the visible color space.



Computer vision and image processing

Color camera imply a color space approximation, from high spectral color space to lower spectral color sub-spaces.



HSV color model

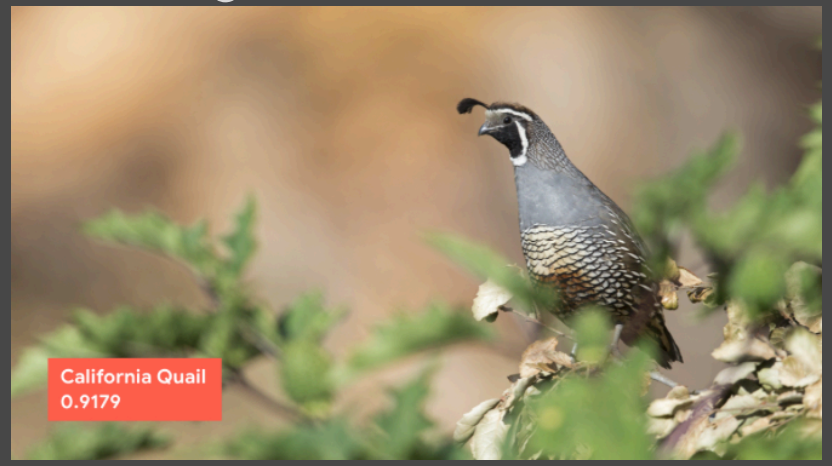
Three channels :

1. Hue (color of the pixel),
2. Saturation (intensity of the color)
3. Value (brightness of the pixel)

Unlike RGB, HSV allows not only to filter image content based on the pixel colors, but also by their intensity and brightness.

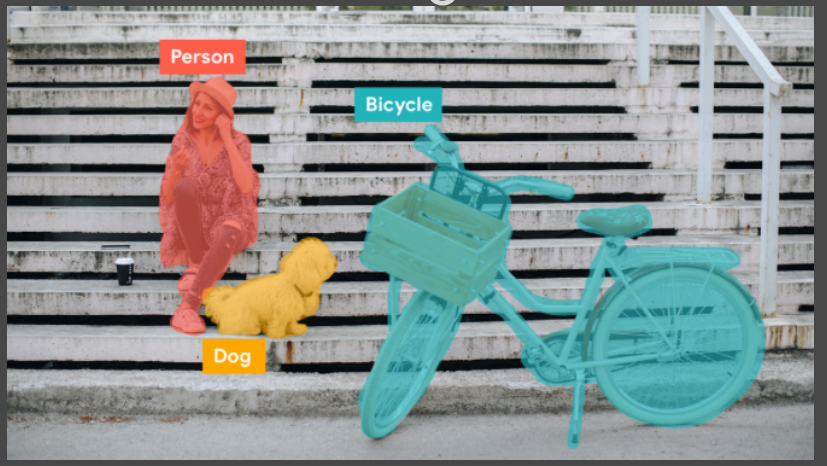
Computer vision and image processing
Pattern recognition (TPU/GPU/CPU)

Image classification



Object detection

Semantic segmentation



Pose estimation

From 2D to 3D Pattern recognition…

Sources
https://coral.ai/models/

Computer vision and image processing

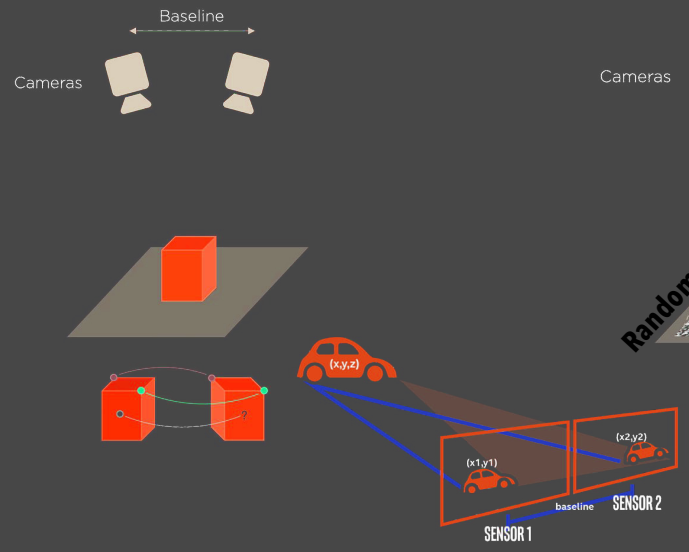
Depth measurement

The ability to determine 3D information about the environment.

Space domain approaches

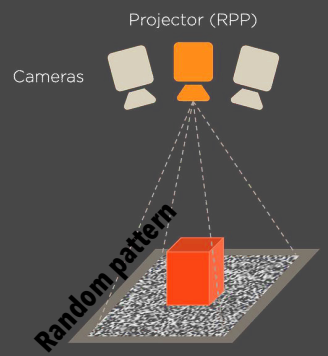
Time domain approaches

PASSIVE STEREO



Depends on the available light in the environment.
Needs textured surfaces.

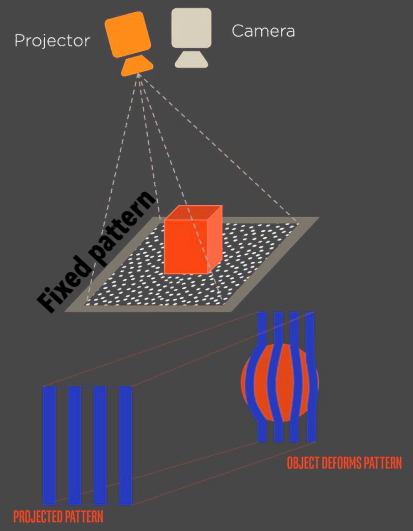
ACTIVE STEREO



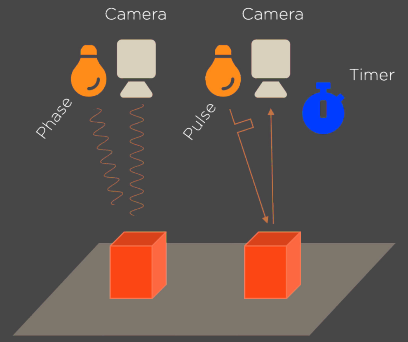
Performs well in the non-textured objects (thanks to the projected pattern).
Unlike structured light depth camera, active stereo depth camera performs better outdoor.
Performs well in low light (thanks to IR projector).

Performs poorly on shiny and absorptive surfaces

STRUCTURED LIGHT



TIME OF FLIGHT



Pixel level processing, very fast/accurate

Works well outdoor but performs poorly on high reflection surfaces, is ineffective during heavy rain…

Sources

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

Computer vision and image processing

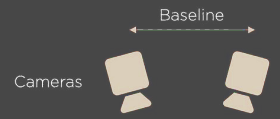
Depth measurement

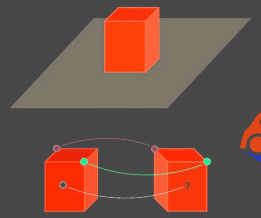
The ability to determine 3D information about the environment.

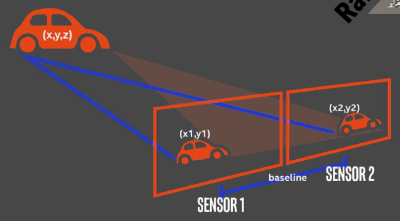
Space domain approaches

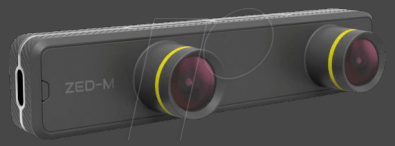
Time domain approaches

PASSIVE STEREO

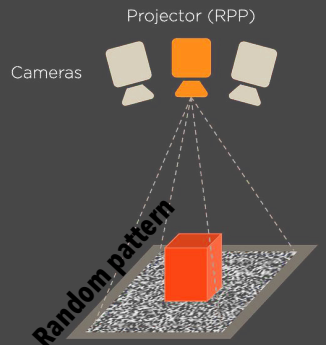


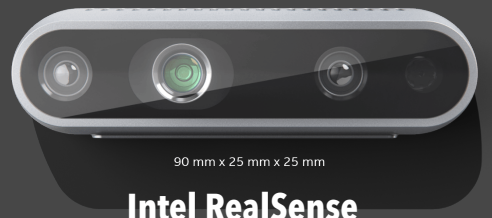






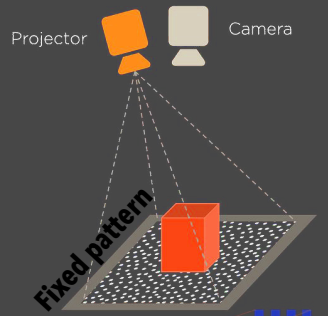
ACTIVE STEREO

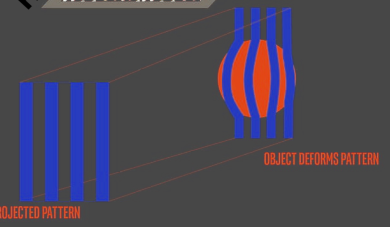




Intel RealSense

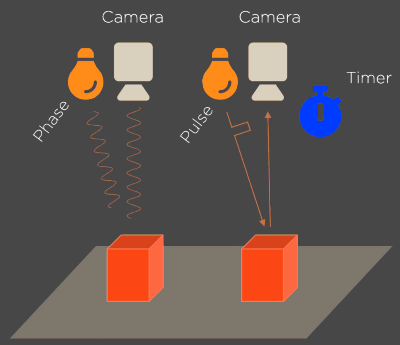
STRUCTURED LIGHT

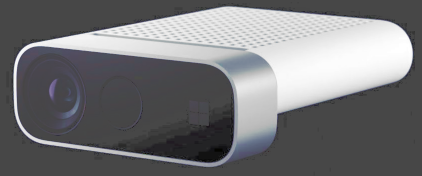




Microsoft Kinect v1

TIME OF FLIGHT

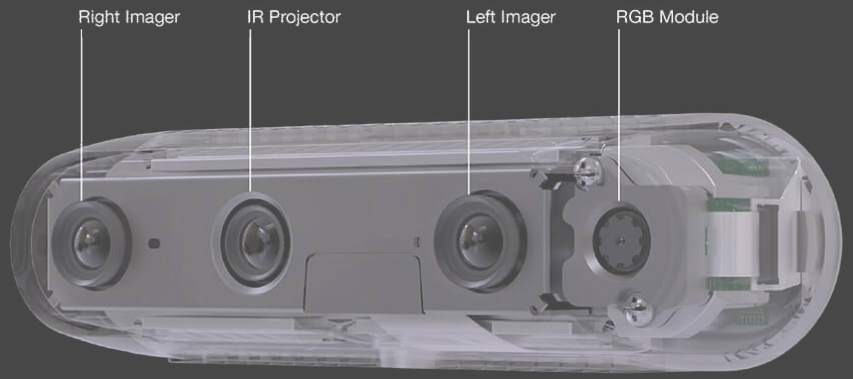




Azure Kinect

Computer vision and image processing
Depth measurement

Focus on the Intel d435**i** depth camera

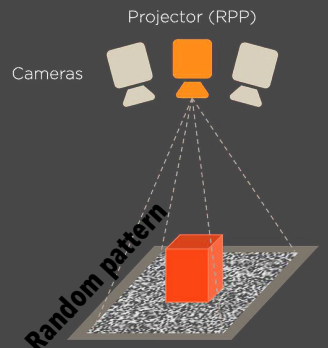


IMU sensor embedded

Provides extrinsics rotation matrix

Space domain approach

ACTIVE STEREO



Sources

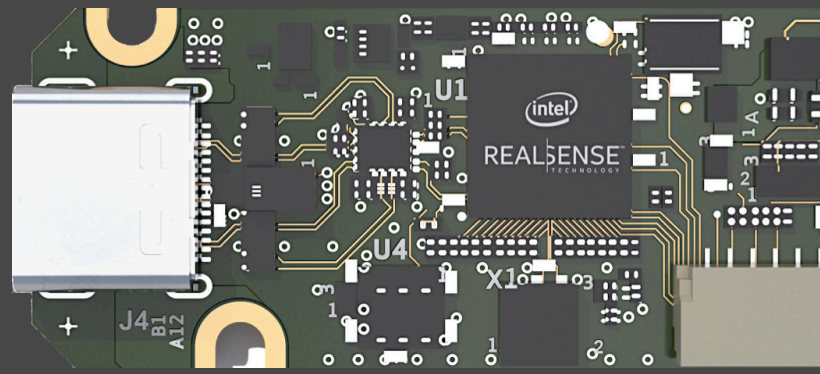
https://www.intelrealsense.com/beginners-guide-to-depth/

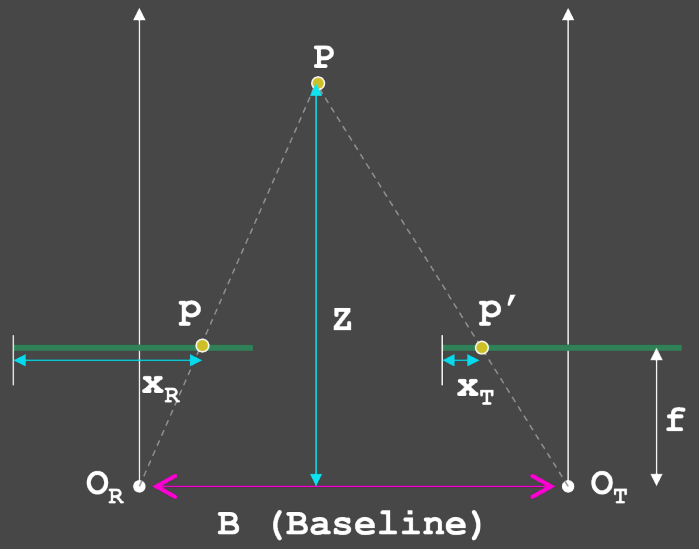
Computer vision and image processing
Depth measurement (on-board VPU)

Intel D435i D400 vision processor on-board computations

Advanced stereo algorithms such as Semi Global matching,

Image rectification for camera optics and alignment compensation



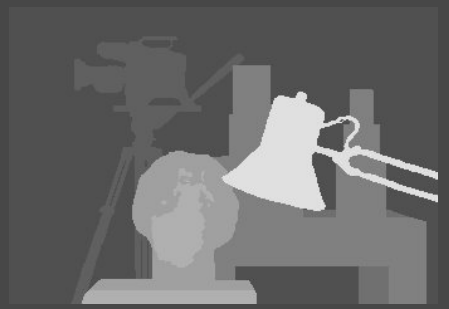


$$\frac{B}{Z} = \frac{(B + x_T) - x_R}{Z - f}$$

$$Z = \frac{Bf}{x_R - x_T}$$

Disparity

→



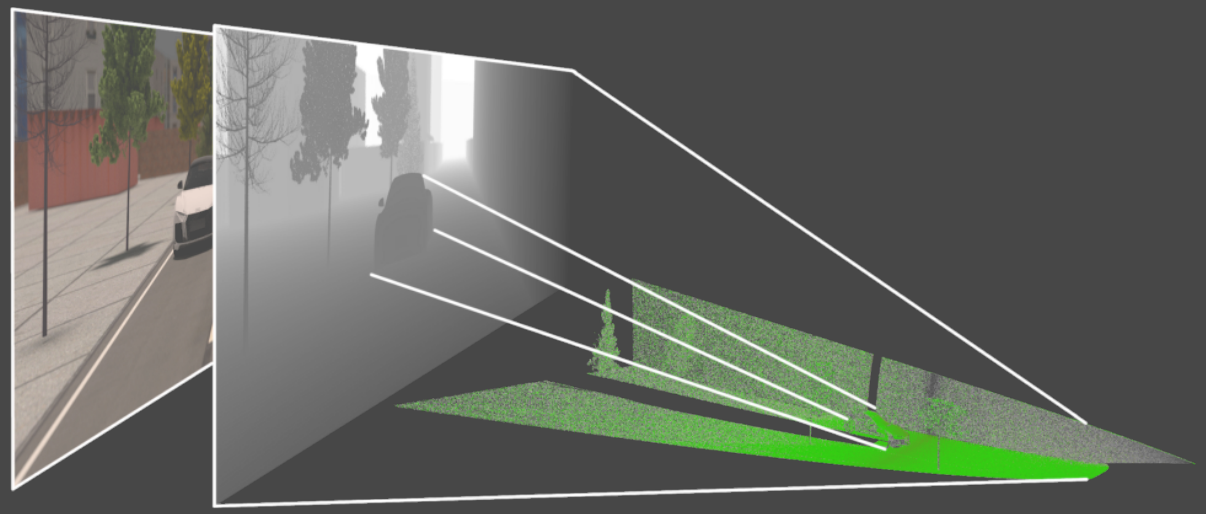
Disparity map

Sources

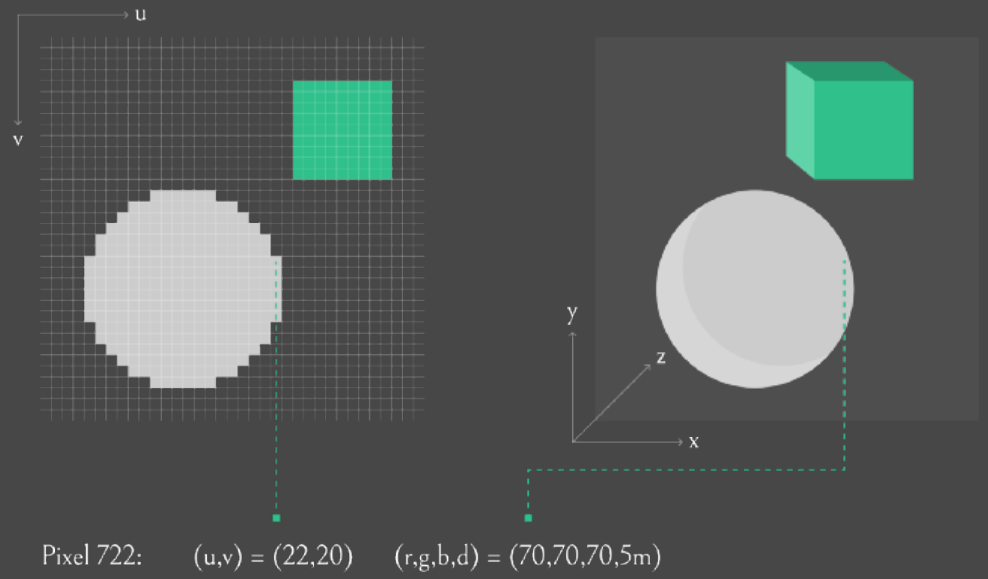
http://vision.deis.unibo.it/~smatt/Seminars/StereoVision.pdf https://www.intelrealsense.com/beginners-guide-to-depth/

Computer vision and image processing
Depth measurement (on-board VPU)

The Intel d435i camera provides a 2D map where each pixel value is the distance from the camera (color camera + depth → RGBD).



https://medium.com/yodayoda/from-depth-map-to-point-cloud-7473721d3f



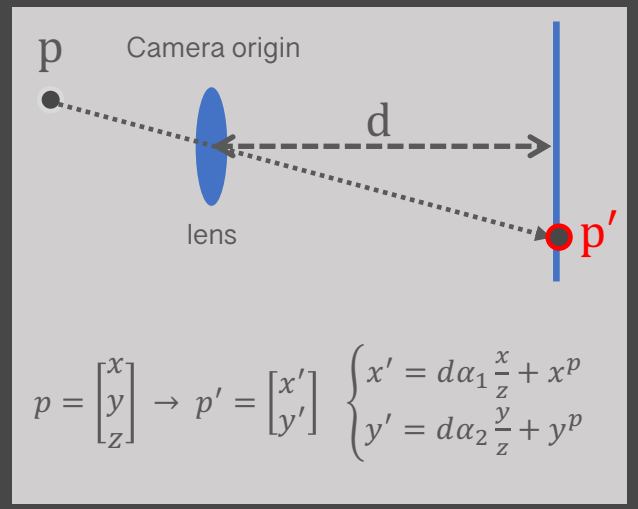
https://medium.com/yodayoda/from-depth-map-to-point-cloud-7473721d3f

On the d435 camera, the depth value is encoded on 16 bits. Value in meter is given by

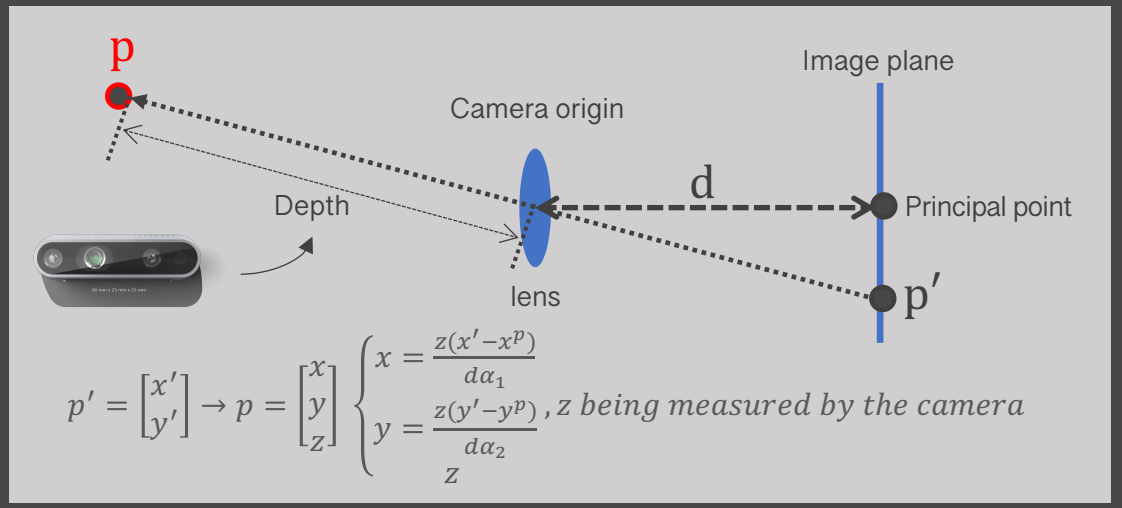
$$Z_{meter} = \frac{depth{-}map\ pixel\ value}{depth{-}scale}$$

Computer vision and image processing
3D space reconstruction (CPU/GPU)

Back-projection from 2D pixel sensor coordinates to 3D camera coordinate system (i.e., without applying extrinsics parameters)



3D → 2D

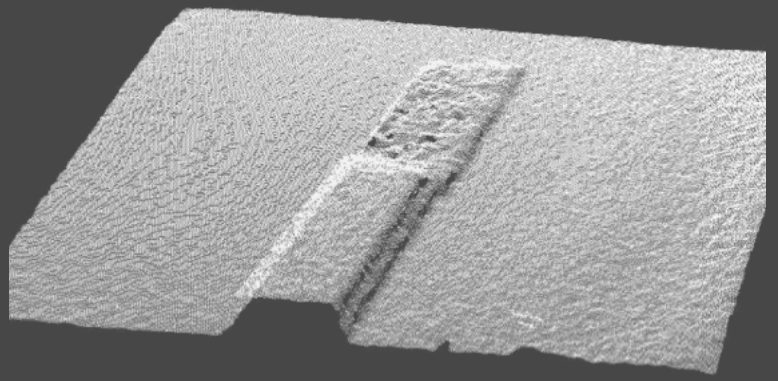


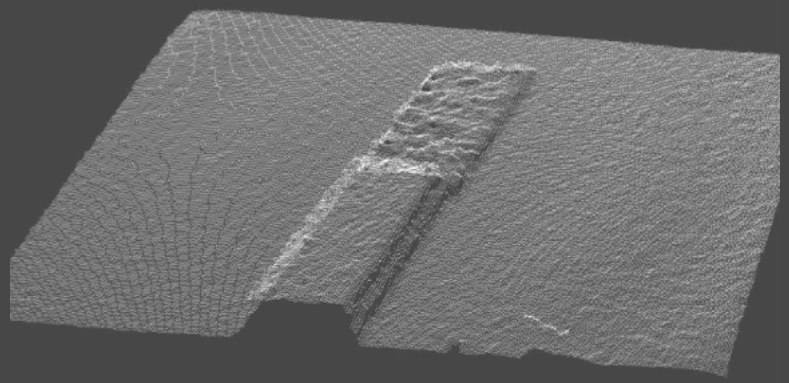
2D → 3D

Computer vision and image processing
Post-processing filtering (CPU/GPU)

Decimation filter

Reduce depth scene complexity by applying kernels ([2x2],[8x8]). Resulting depth map is scaled down, each pixel depth value corresponding to the mean kernel value.



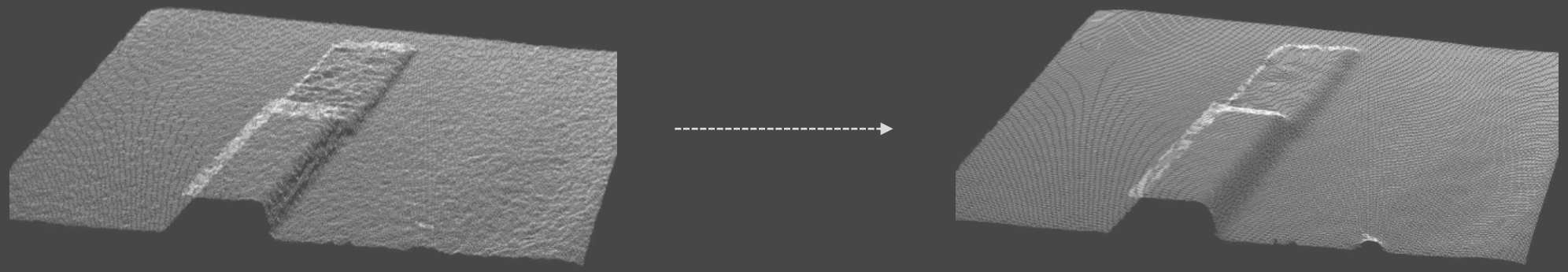


https://www.intel.com/content/dam/support/us/en/documents/emerging-technologies/intel-realsense-technology/Intel-RealSense-Depth-PostProcess.pdf

Computer vision and image processing
Post-processing filtering (CPU/GPU)

Spatial (edge-preserving) filter

The filter performs a series of 1D horizontal and vertical passes or iterations, to enhance the smoothness of the reconstructed data.

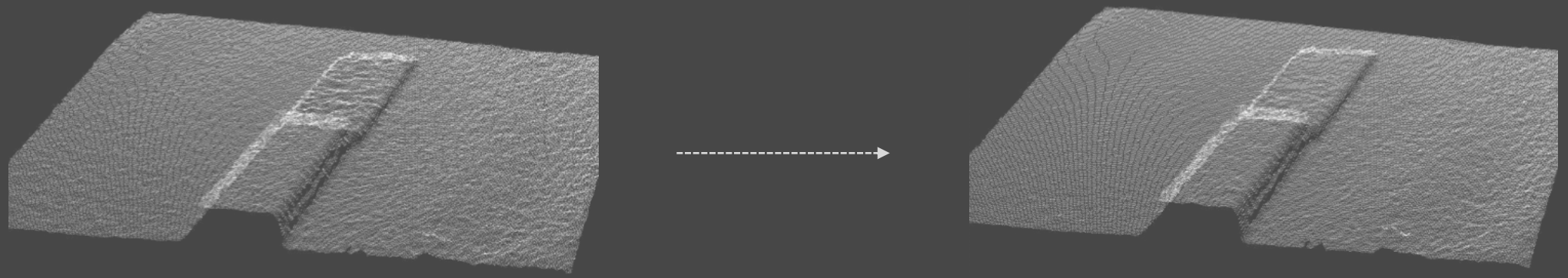


https://www.intel.com/content/dam/support/us/en/documents/emerging-technologies/intel-realsense-technology/Intel-RealSense-Depth-PostProcess.pdf

Computer vision and image processing
Post-processing filtering (CPU/GPU)

Temporal filter

The temporal filter is intended to improve the depth data persistency by manipulating per-pixel values based on previous frames.



Persistency on moving camera/objects ?

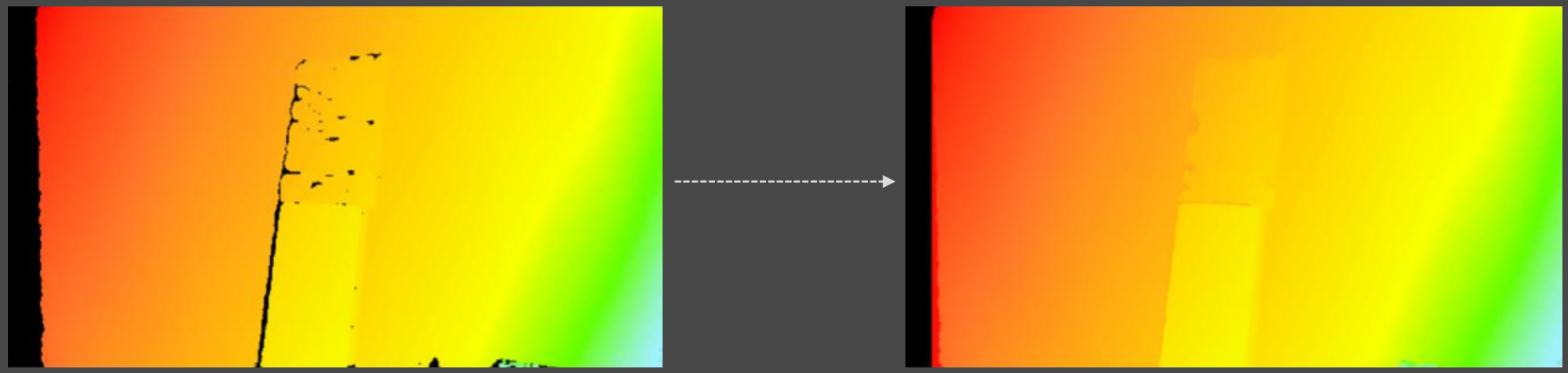
https://www.intel.com/content/dam/support/us/en/documents/emerging-technologies/intel-realsense-technology/Intel-RealSense-Depth-PostProcess.pdf

Computer vision and image processing
Post-processing filtering (CPU/GPU)

Spatial holes filling filter

The filter implements several methods to rectify missing data (depth=0) in the resulting image. Holes in depth map result from:

1) **Occlusion** – the left and right images do not see the same object due to shadowing.
2) **Lack of texture**, light absorption, shininess
3) **Multiple matches**
4) **No signal** – under/over image exposition
5) **Below min z**



https://www.intel.com/content/dam/support/us/en/documents/emerging-technologies/intel-realsense-technology/Intel-RealSense-Depth-PostProcess.pdf

Computer vision and image processing
Post-processing filtering (CPU/GPU)

Filtering does not come for free !!!

1280 x 720 resolution, i7-6950X

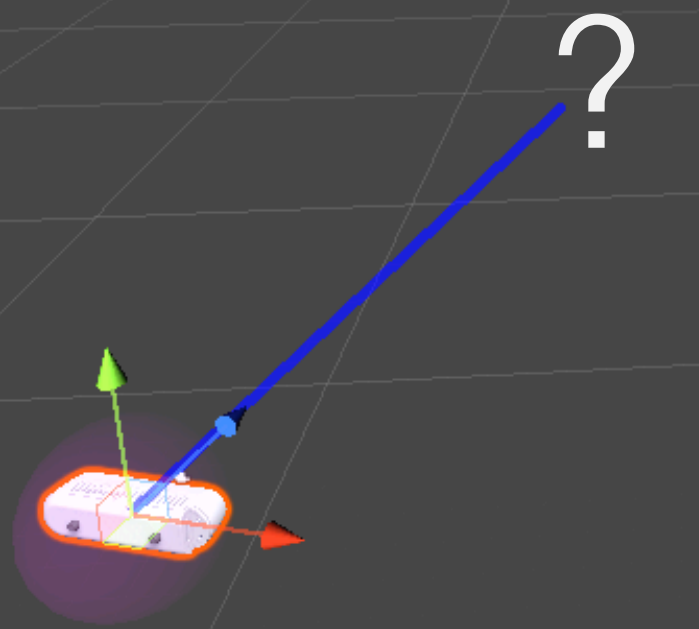
Sub-sample factor	Subsample Time, ms	Depth-to-Disparity Time, ms	Spatial Filtering Time, ms (+ hole filling*)	Temporal Filtering Time, ms	Disparity-to-Depth Time, ms	Hole Filling Time, ms	Total Processing Time, ms (+ Spatial hole filling)
1	0	4.09	22.18 (+ 4.38)	3.13	3.49	3.09	**36.96 (+ 4.38)**
2	4.84	1.83	8.00 (+1.31)	1.32	1.41	1.17	**19.24 (+1.31)**
3	5.81	0.95	3.92 (+1.11)	0.76	0.83	0.68	**13.22 (+1.11)**
4	3.02	0.62	2.61 (+0.56)	0.53	0.57	0.47	**7.97 (+0.56)**
5	2.91	0.44	2.01 (+0.54)	0.39	0.40	0.33	**6.60 (+0.54)**
6	2.85	0.34	1.34 (+0.39)	0.30	0.30	0.26	**5.47 (+0.39)**

Hole radius set to 2

https://www.intel.com/content/dam/support/us/en/documents/emerging-technologies/intel-realsense-technology/Intel-RealSense-Depth-PostProcess.pdf

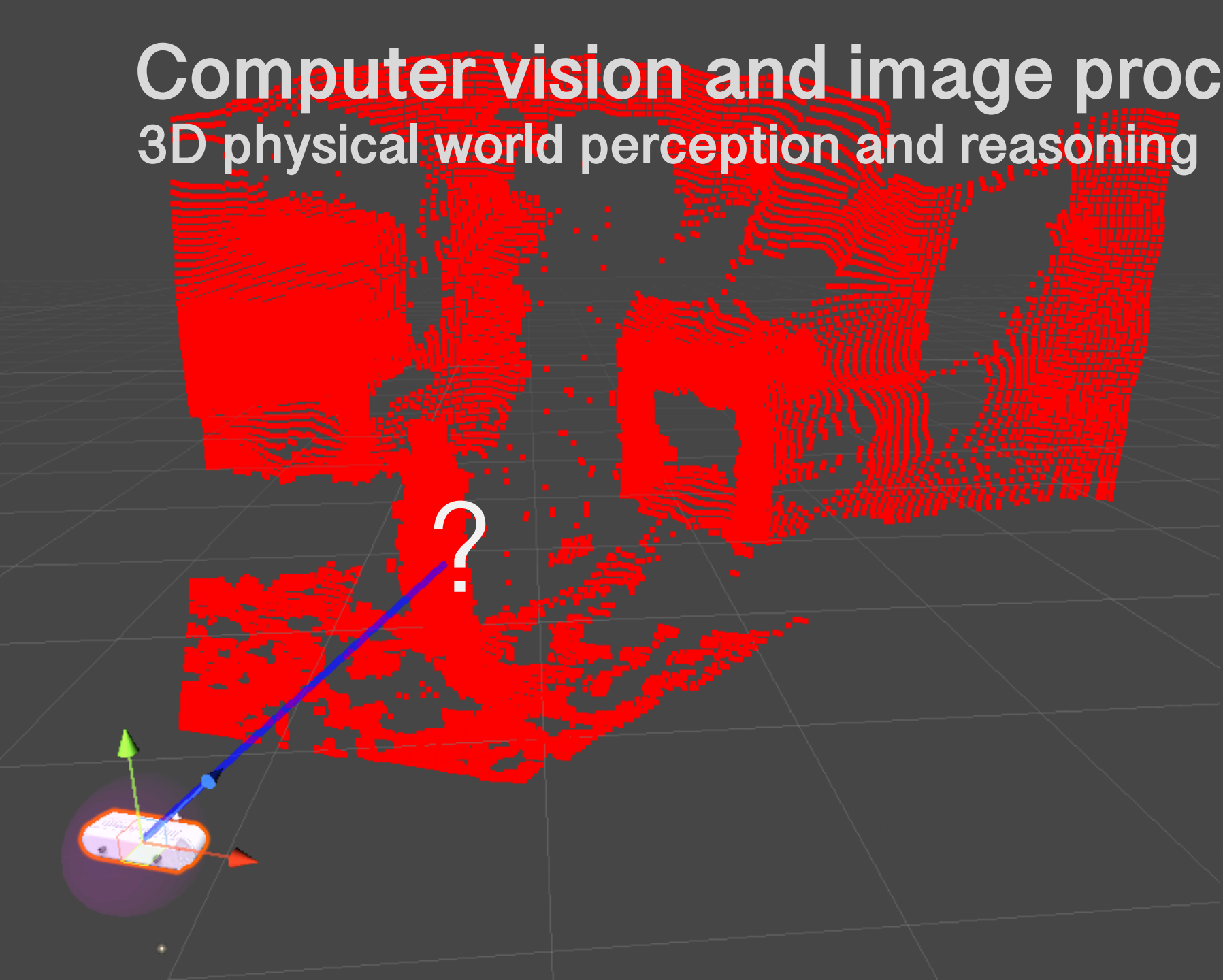
Computer vision and image processing

Distance measure (1D)



Computer vision and image processing

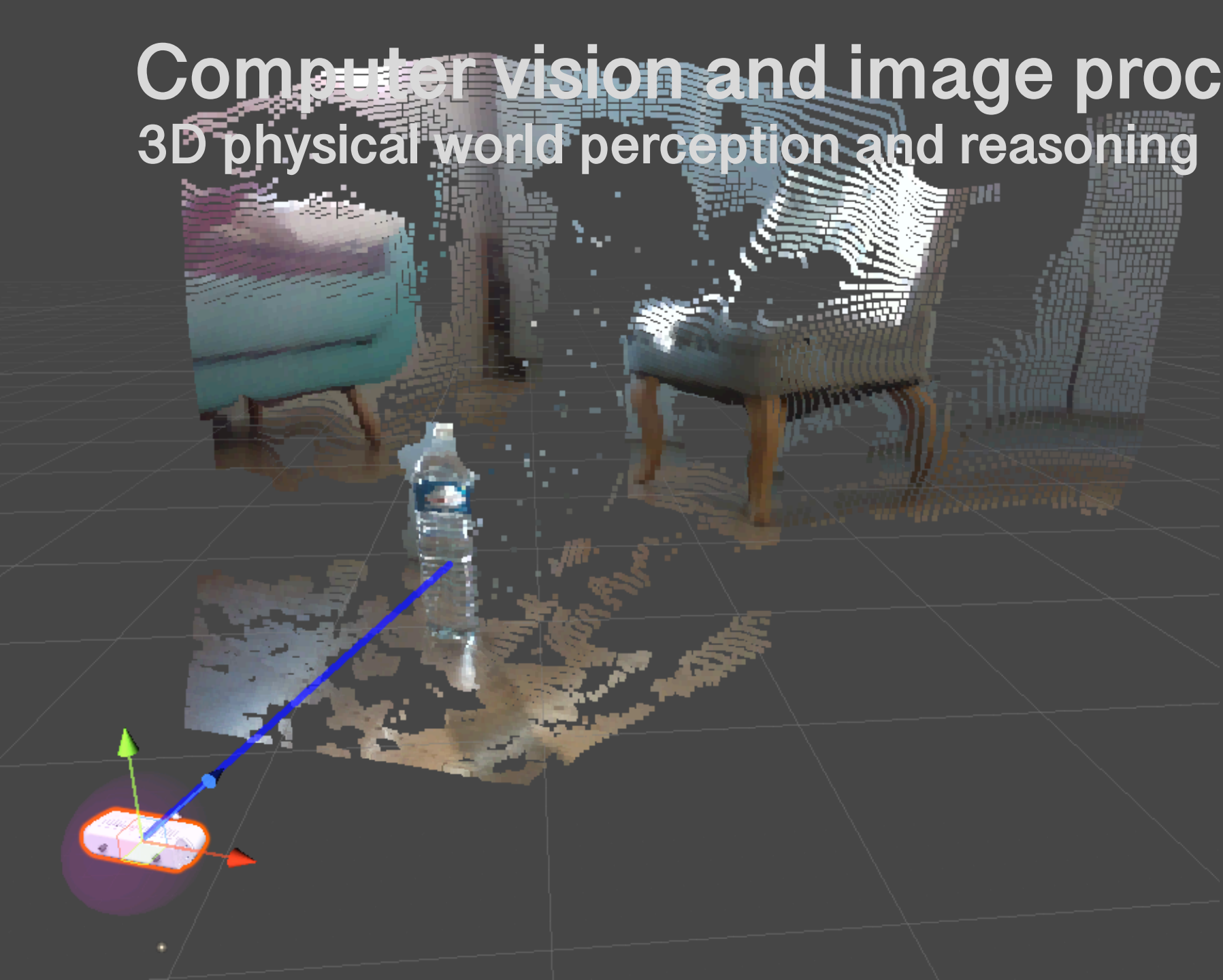
3D physical world perception and reasoning



Point cloud (3D)

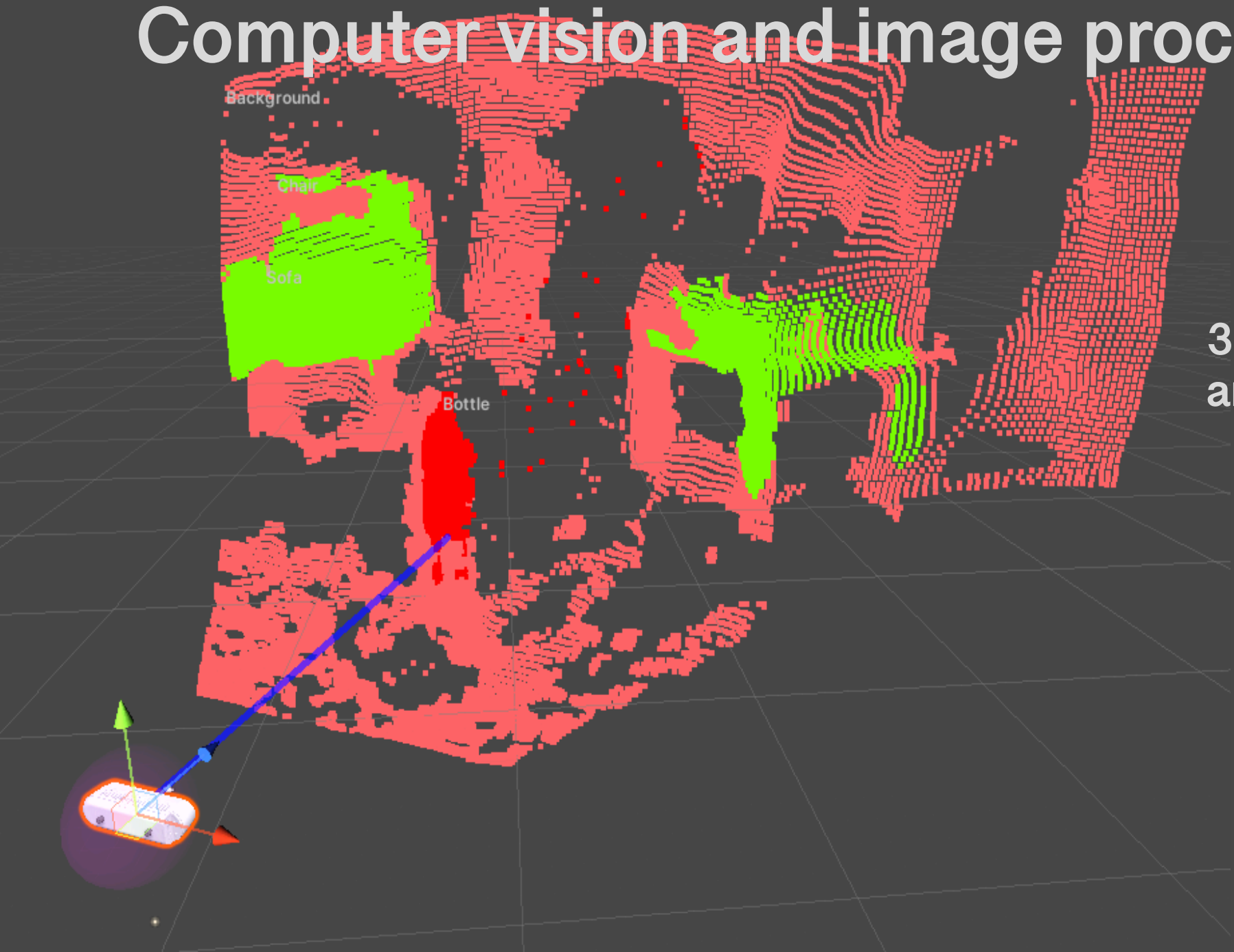
Computer vision and image processing

3D physical world perception and reasoning



3D reconstruction

Computer vision and image processing

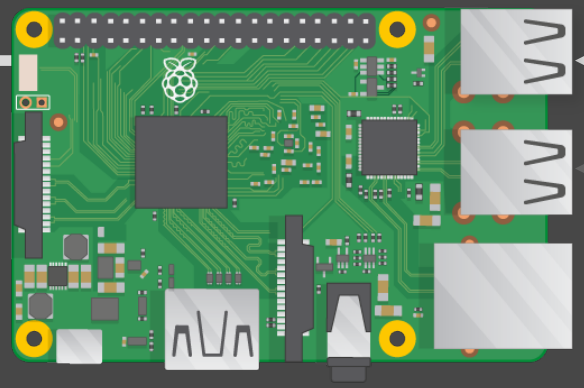


3D semantic segmentation and object recognition

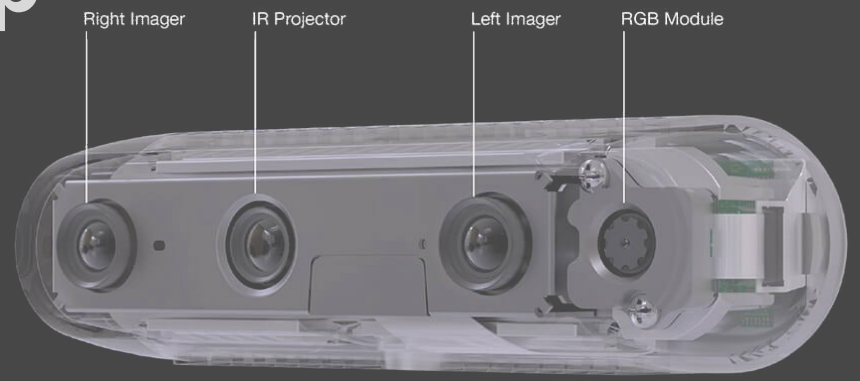
AlphaBot2 Pi visual perception setup

RaspberryPi

WiFi

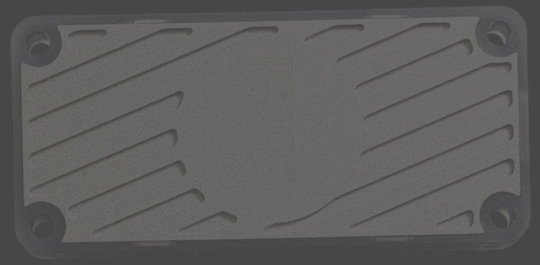


USB



Intel Realsense D435i

Vision processor V4 (VPU)
Inertial Measurement Unit (IMU)
Stereoscopic depth (1280x720 90fps)
2MP RGB (1920 x 1080 30fps)



Google CORAL EdgeTPU
4 TOPS (INT8)

RaspberryPi 3B+

Broadcom BCM2837B0, quad-core Cortex-A53 (ARMv8) 64-bit SoC @ 1.4GHz
Broadcom Videocore-IV GPU
1Gb LPDDR2 SDRAM
2.4GHz and 5GHz IEEE 802.11.b/g/n/ac wireless LAN
4 x USB2.0 ports

RaspberryPi 4B

Broadcom BCM2711, quad-core Cortex-A72 (ARMv8) 64-bit SoC @ 1.5GHz
Broadcom Videocore-VI GPU
1Gb LPDDR4 SDRAM
2.4GHz and 5GHz IEEE 802.11.b/g/n/ac wireless LAN
2 x USB2.0 & 2 x USB3.0 ports

Sources and references

Intelligent systems

[1] https://www.informatik.uni-bonn.de/en/research-and-phd/intelligent-systems

[2] Autonomous Mobile Robots Roland Siegwart, Margarita Chli, Nick Lawrance

Camera model

[3] https://cvgl.stanford.edu/teaching/cs231a_winter1415/lecture/lecture2_camera_models_note.pdf

[4] Cyrill Stachniss, cyrill.stachniss@igg.uni-bonn.de, https://www.ipb.uni-bonn.de/html/teaching/photo12-2021/2021-pho1-20-camera-params.pptx.pdf

Stereo vision

[5] http://vision.deis.unibo.it/~smatt/Seminars/StereoVision.pdf

Intel D435i depth camera

[9] https://dev.intelrealsense.com/docs/whitepapers#section-d400

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

[8] https://www.intel.com/content/dam/support/us/en/documents/emerging-technologies/intel-realsense-technology/Intel-RealSense-Depth-PostProcess.pdf