KENTECH
Korea Institute of Energy Technology

Visionary Course – Energy AI
Week 10

Seokju Lee

KENTECH
Korea Institute of Energy Technology

Semantic Segmentation

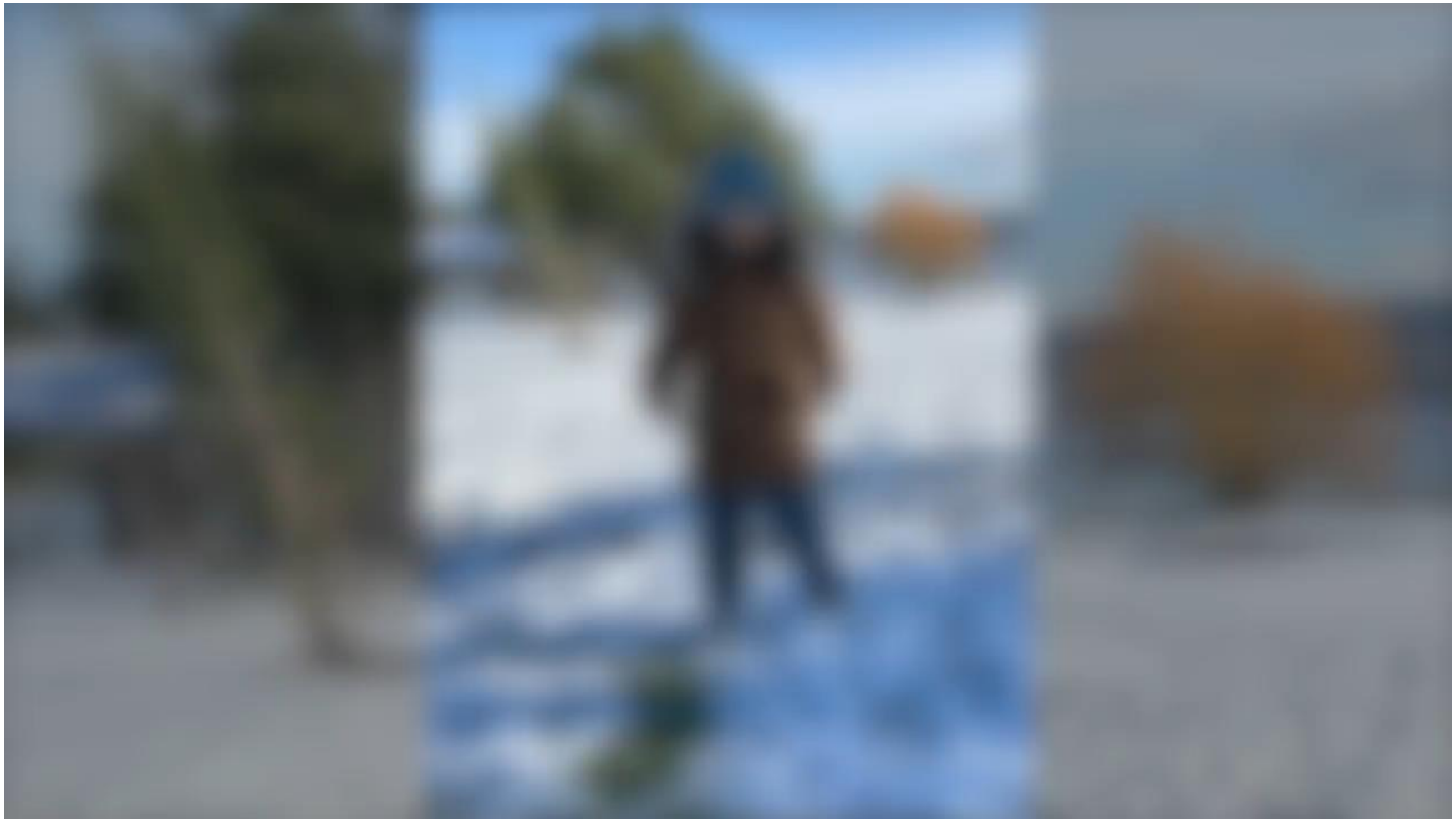
Photo Segmentation on My iPhone

In the newest version of **iPhone** (iOS 16),

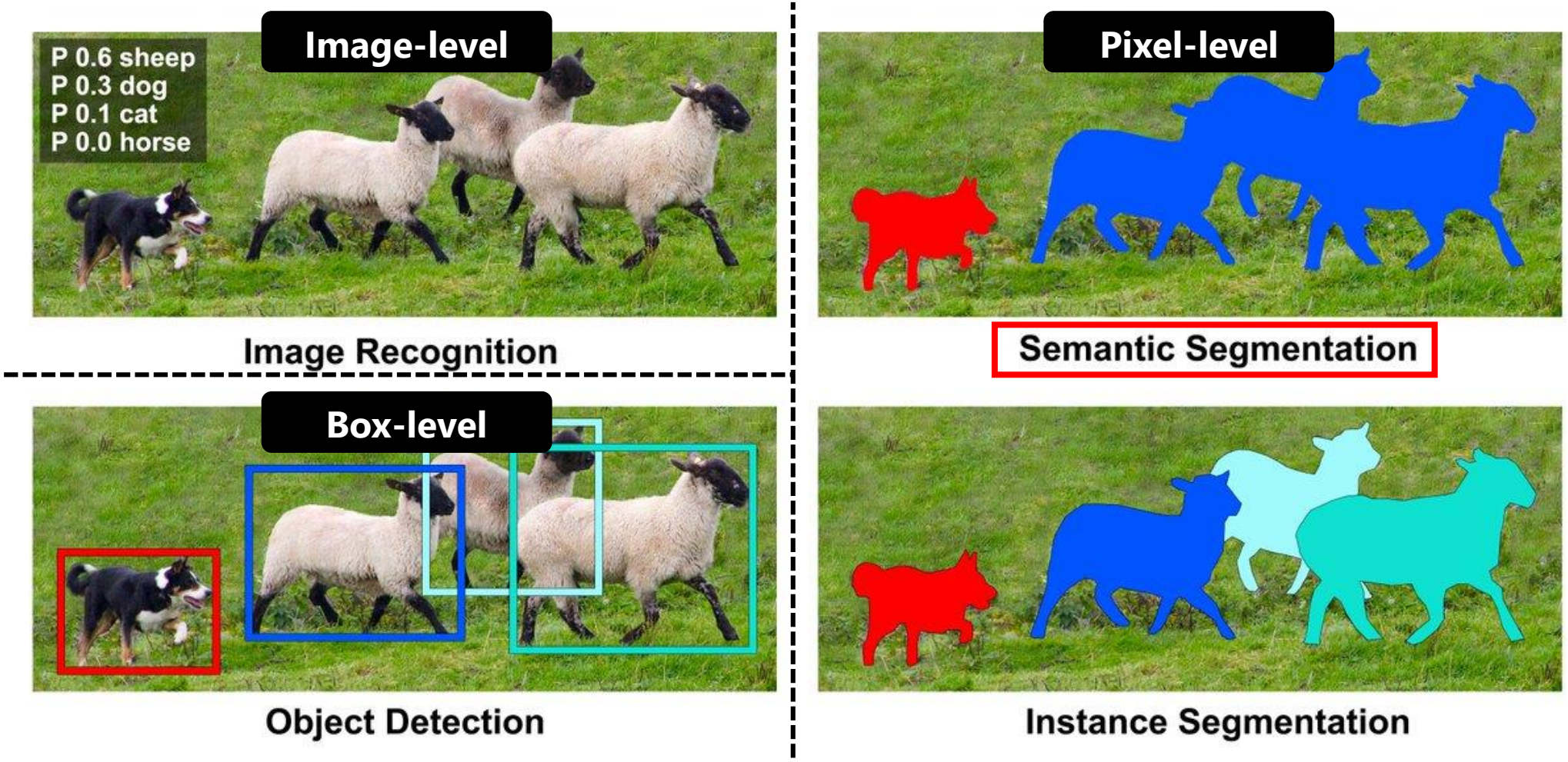
Image segmentation!

Image Segmentation + 3D Computer Vision

3D Photography is available,



Computer Vision Tasks



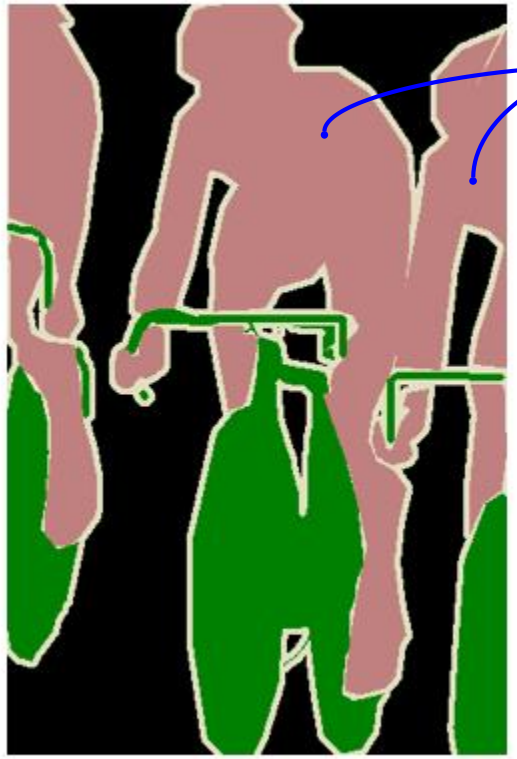
[1] Figure credits: John Wilson

Semantic Segmentation

: Task of assigning **a label** for **every pixel** of an image with a corresponding **class**

Dense prediction

predict



Different objects,
but same prediction
→ limitation

Person
Bicycle
Background

[1] Figures from Pascal VOC Challenge 2012

Semantic Segmentation: Input & Output

Groud Truth (GT) → supervised learning

segmented

1: Person
2: Purse
3: Plants/Grass
4: Sidewalk
5: Building/Structures

3	3	3	3	3	3	3	3	3	3	3	3	5	5	5	5	5	5
3	3	3	3	3	3	3	3	3	3	3	3	5	5	5	5	5	5
3	3	3	3	3	3	1	1	3	3	3	3	5	5	5	5	5	5
3	3	3	3	3	1	1	1	1	3	3	3	5	5	5	5	5	5
3	3	3	3	3	3	1	1	3	3	3	5	5	5	5	5	5	5
5	5	3	3	3	3	1	1	3	3	5	5	5	5	5	5	5	5
4	4	3	4	1	1	1	1	1	1	4	4	4	5	5	5	5	5
4	4	3	4	1	1	1	1	1	1	4	4	4	4	4	5	5	5
4	4	4	1	1	1	1	1	1	1	1	4	4	4	4	4	4	4
3	3	3	1	1	1	1	1	1	1	1	4	4	4	4	4	4	4
3	3	3	1	2	2	1	1	1	1	1	4	4	4	4	4	4	4
3	3	3	1	2	2	1	1	1	1	1	4	4	4	4	4	4	4

Input

→ Input image resolution: $\boldsymbol{W} \times \boldsymbol{H} \times \boldsymbol{3}$

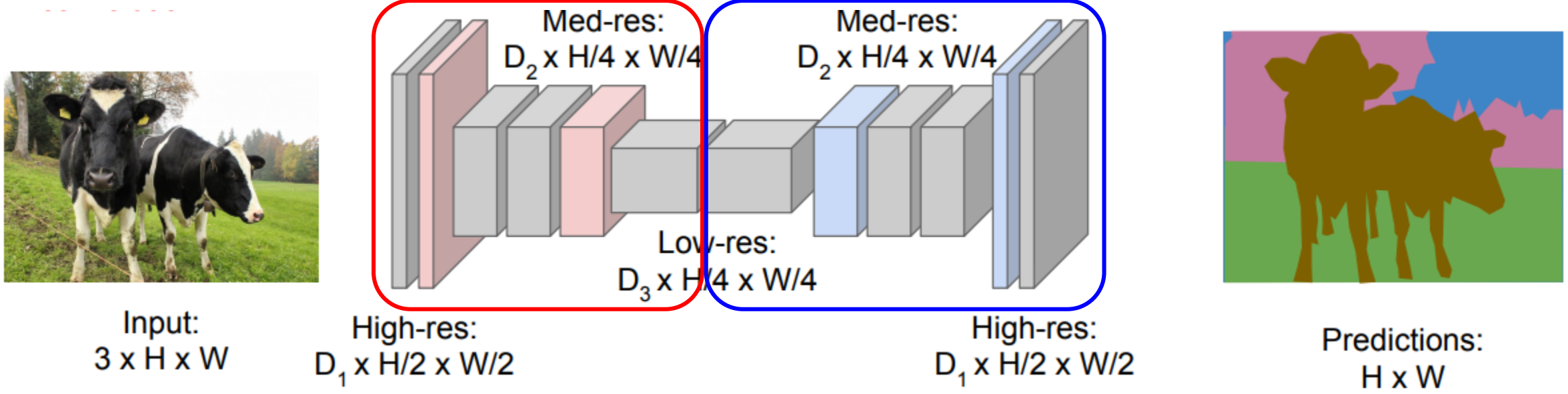
Semantic Labels

Note that this is a low-resolution prediction map for visual clarity.
In reality, the segmentation label resolution should match the original input's resolution ($\boldsymbol{W} \times \boldsymbol{H} \times \boldsymbol{3}$).

[1] Figure credits: Jeremy Jordan

Fully Convolutional Approach

Design a network as a bunch of convolutional layers,
with **downsampling** and **upsampling** inside the network!



Downsampling: **Feature extraction** by pooling, strided convolution (encoder)

Upsampling: **Recover of spatial resolution** by unpooling, strided transpose convolution (decoder)

[1] Stanford university cs231n lecture slide

KENTECH
Korea Institute of Energy Technology

Semantic Segmentation on Jetson Nano

Experiments – Semantic Segmentation

Live semantic segmentation with visualization
*Your basic workspace is here: "**cd ~/jetson-inference/build/aarch64/bin**"

Q1. Run "**python segnet.py --flip-method=rotate-180**". What is on the screen? What is different from detectnet in the previous class (object detection)? Could you guess the meaning of the different colors on the screen?

Q2. Go to the linked page (https://github.com/dusty-nv/jetson-inference/blob/master/docs/segnet-console-2.md). Please read the list of segmentation models. Try to use different models, trained on different datasets (upto your preference). For example, run "**python segnet.py --network=fcn-resnet18-cityscapes-1024x512 --flip-method=rotate-180**", and let the camera see the driving video.

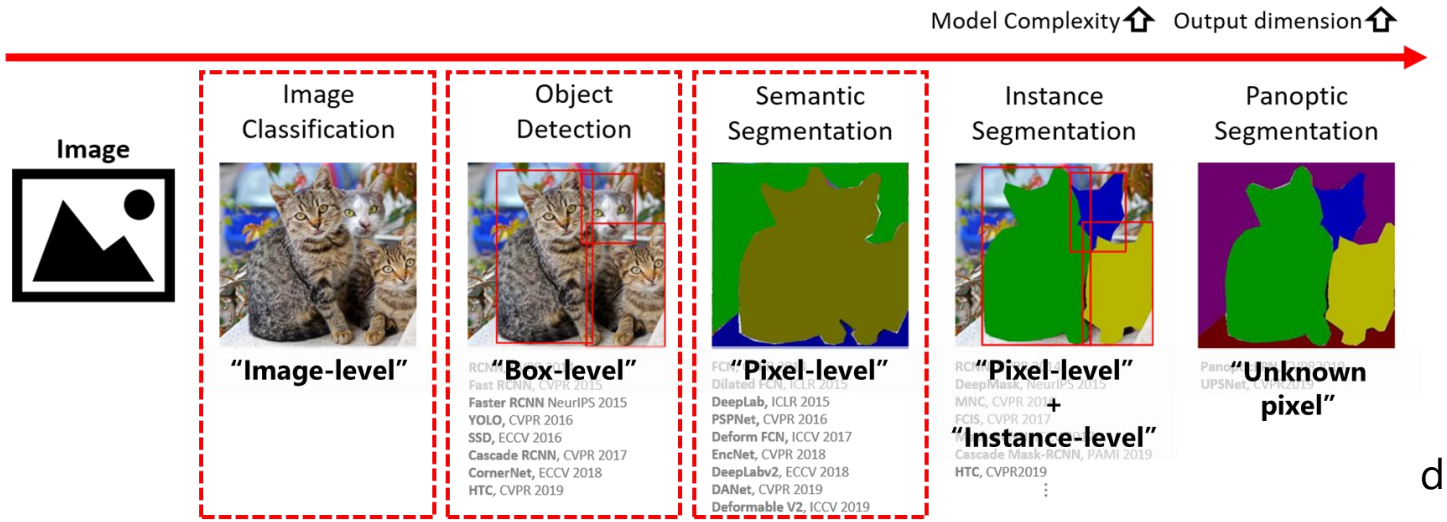
Summary

Basic computer vision tasks (semantic tasks)

→ Image classification (AlexNet, VGG, GoogleNet, ResNet): Basic deep neural networks

→ Object detection (R-CNN, SSD): Trade-off between two-stage & one-stage detectors

→ Semantic segmentation (Fully Convolutional Network): Encoder (feature extraction) + Decoder (upsampling)



→ Combinatorial works of object detection and semantic segmentation

→ We will move on to **geometric tasks**!


Q&A
KENTECH
Korea Institute of Energy Technology

KENTECH
Korea Institute of Energy Technology

3D Visual Perception

Human Visual System

"About **half** of neocortex in humans is devoted to **vision**." [1]

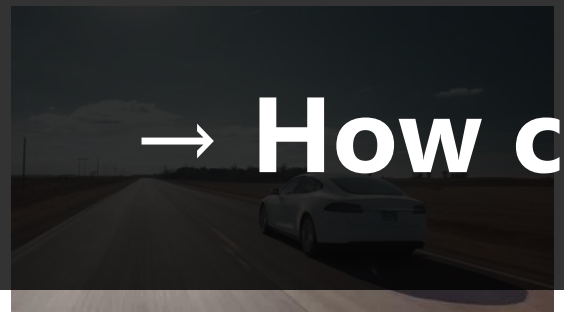
Geometric perception

Where is it?

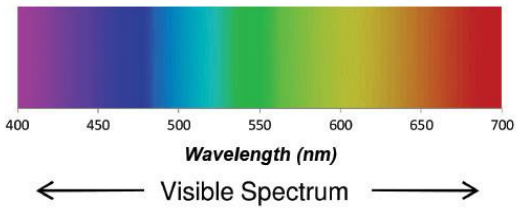
Two independent visual pathways:

"**What** path & **Where** path"

High-level decision



Low-level visual signal



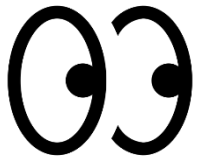
Semantic perception

What is it?

→ **How can we perceive the 3D world?**

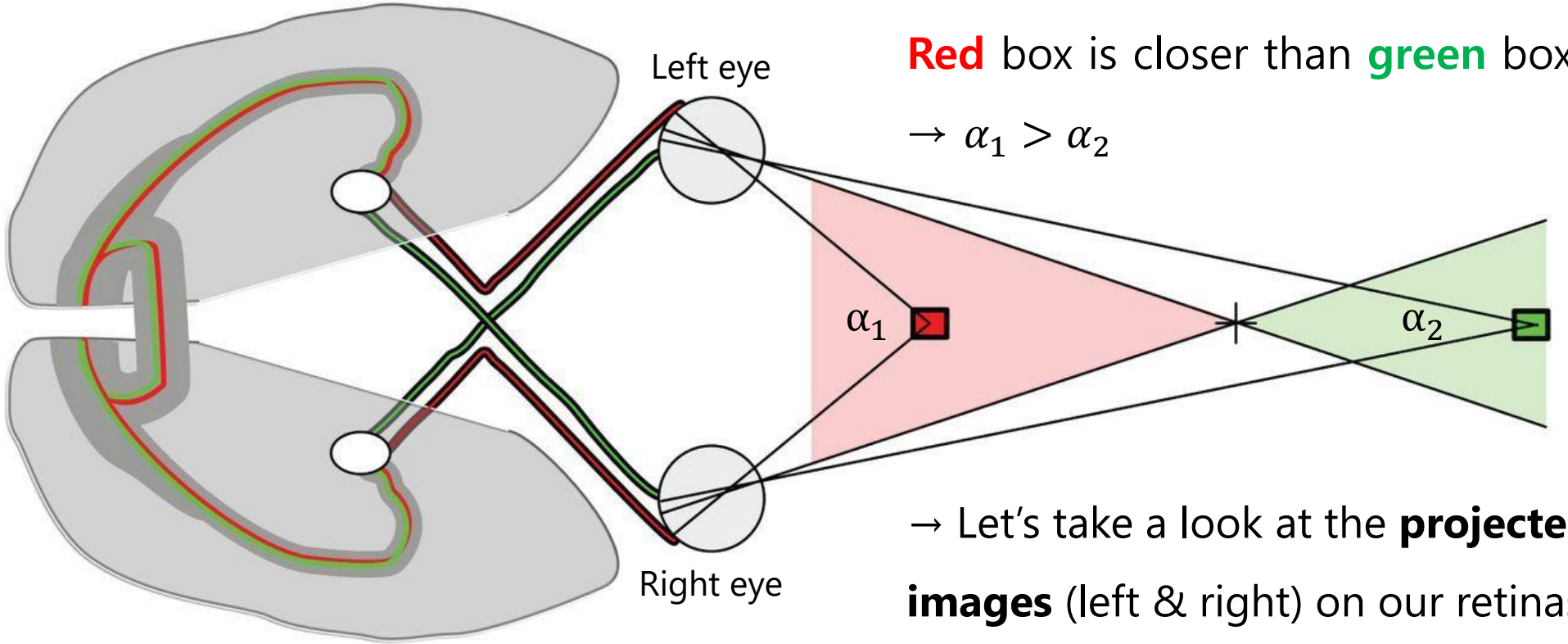
[1] Barton, Robert A. "Visual specialization and brain evolution in primates." *Proceedings of the Royal Society of London* (1998).
[2] M. A. Goodale, et al., "Separate visual pathways for perception and action." *Trends in Neurosciences* (1992).

Human 3D Perception: Binocular System



Binocular correspondence

- Two eyes
- Finding matching points



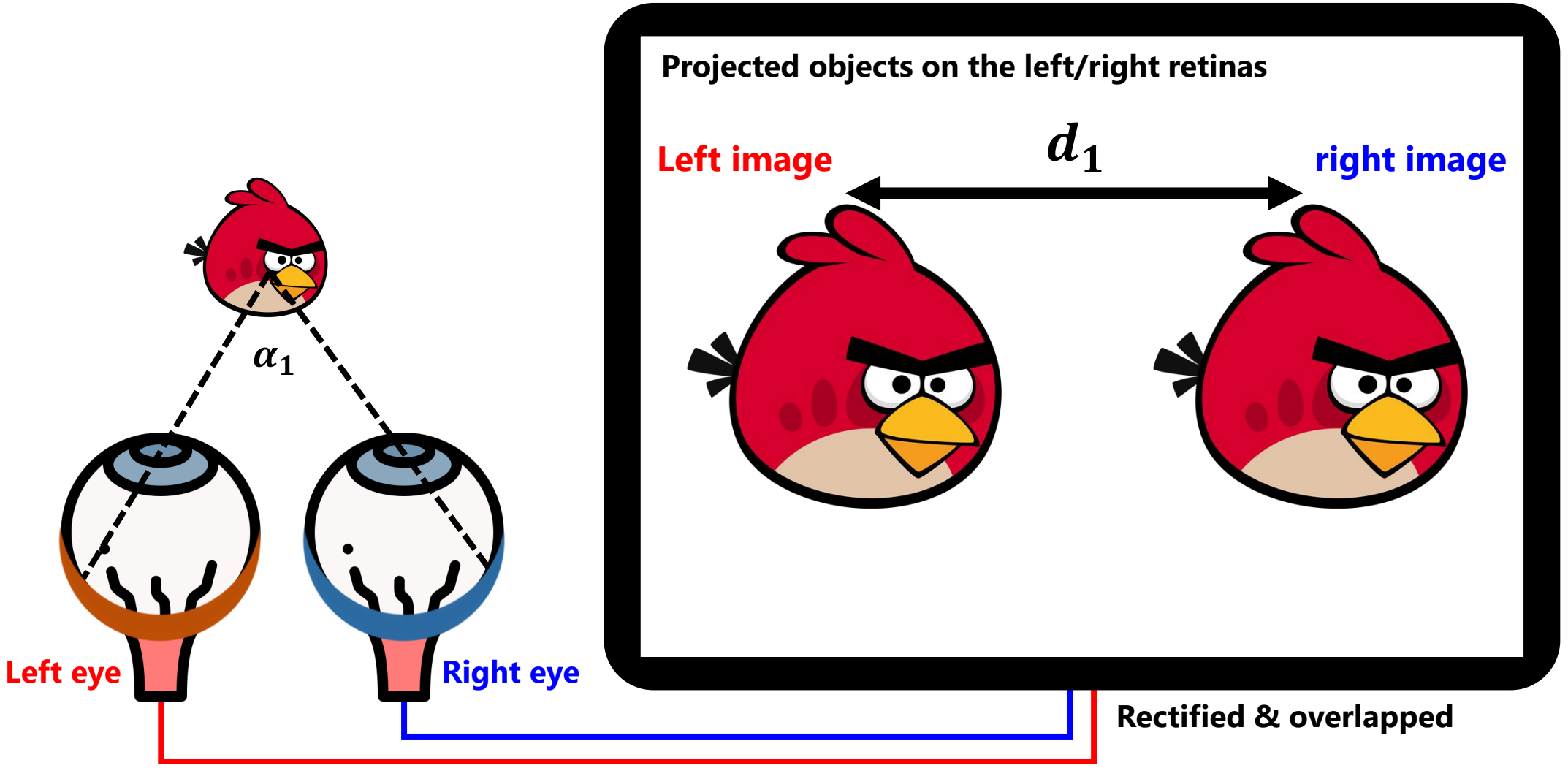
Red box is closer than **green** box

$\rightarrow \alpha_1 > \alpha_2$

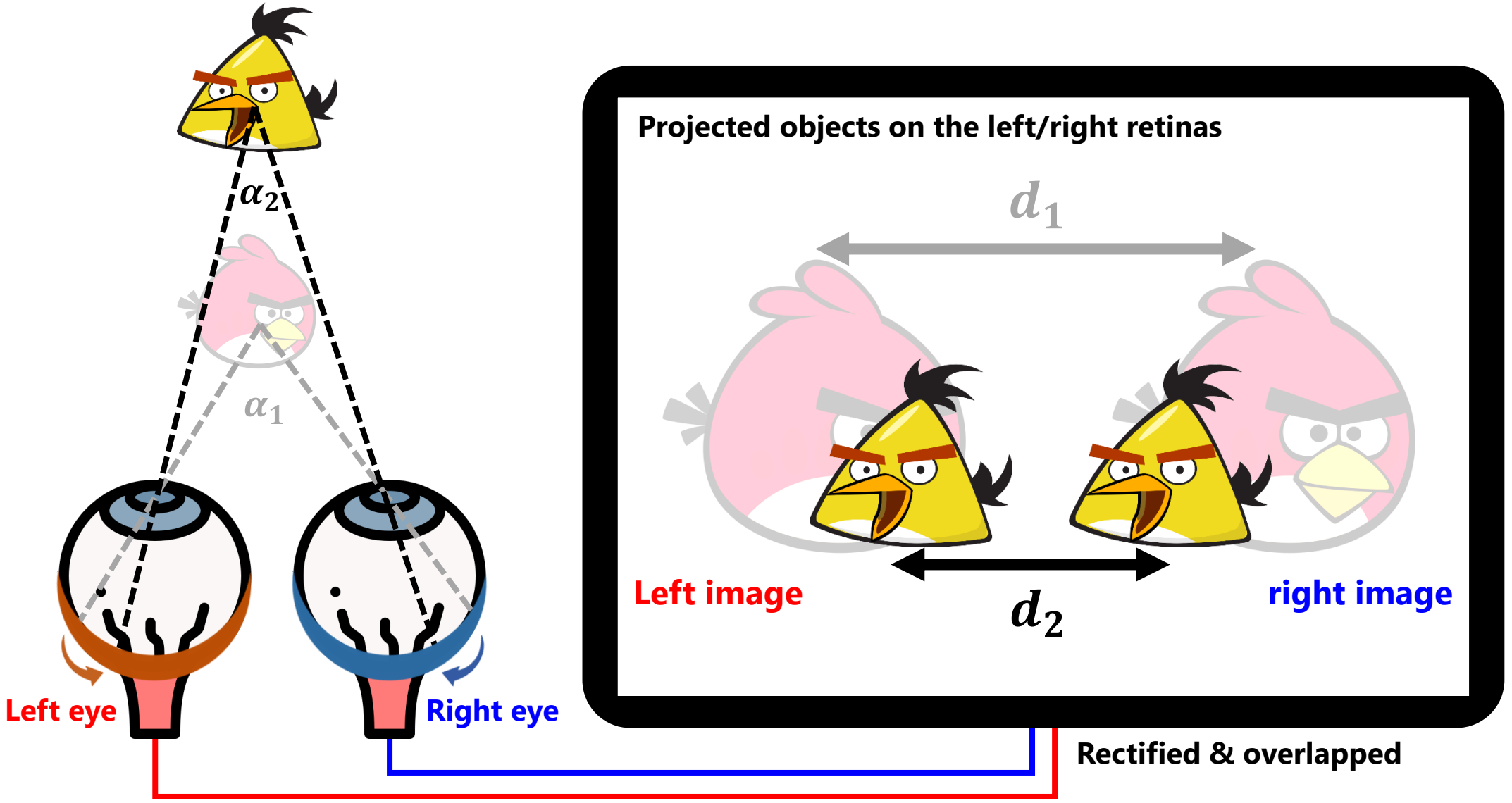
→ Let's take a look at the **projected images** (left & right) on our retinas!

[1] A. Welchman et al., "The Human Brain in Depth: How We See in 3D.", Annual Review of Vision Science 2016.

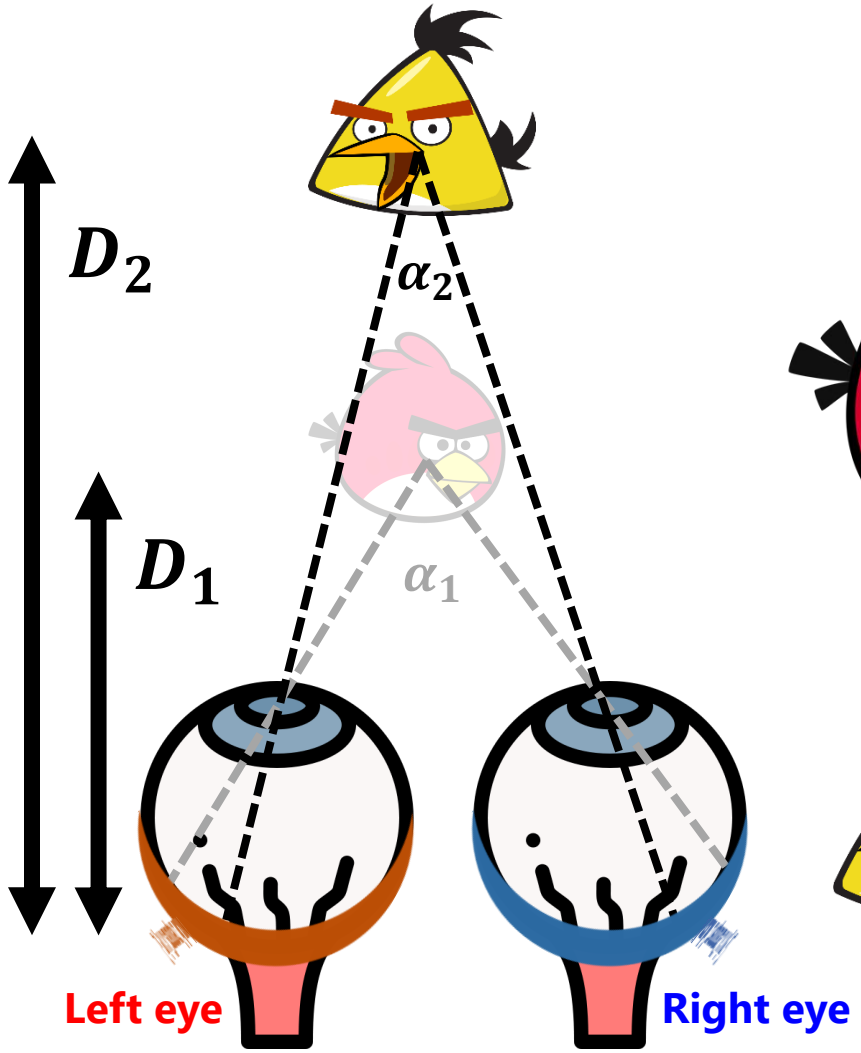
Binocular Correspondence: Near Objects

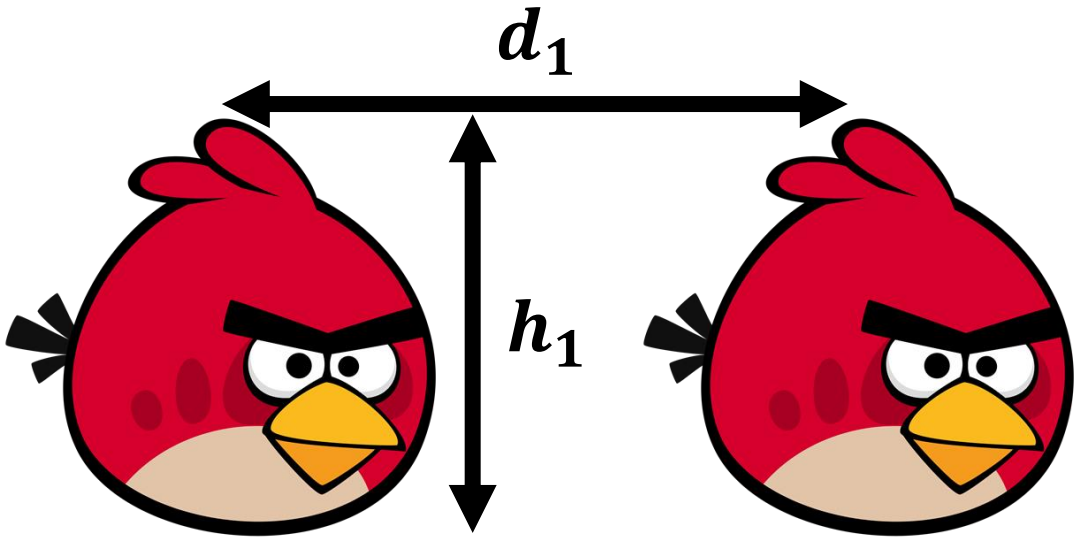


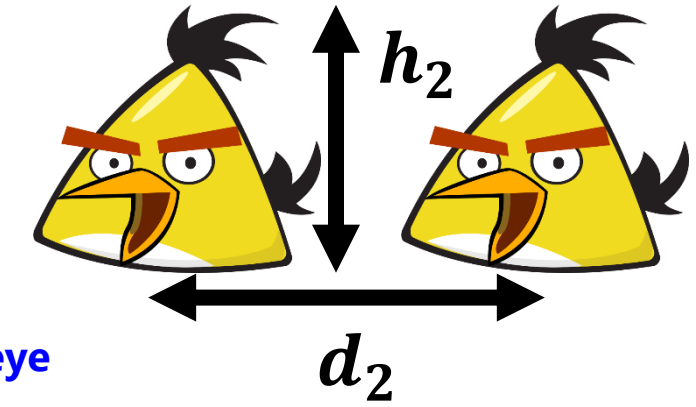
Binocular Correspondence: Far Objects



Binocular Correspondence: Disparity (d)

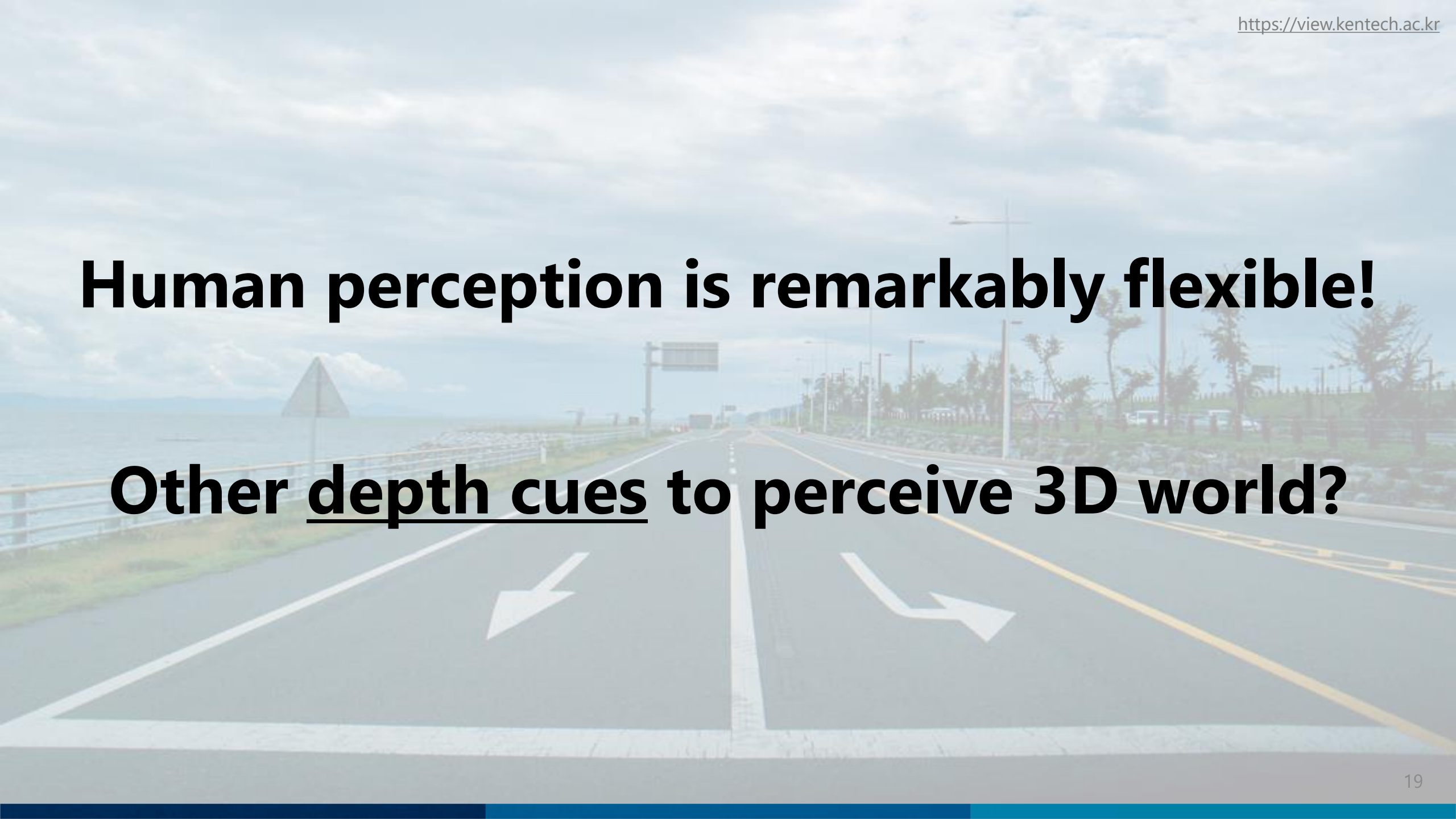






Three important facts

1) Depth (D) ↑ → Angle (α) ↓
2) Depth (D) ↑ → Disparity (d) ↓
3) Depth (D) ↑ → Object's size (h) ↓



Human perception is remarkably flexible!

Other depth cues to perceive 3D world?

Perceiving 3D World with Monocular System

Department of Transportation, USA

Driving With One Eye

Not only is it possible to drive with one eye (assuming that you have good vision in your remaining eye) it's also legal in many states. Though there isn't a federal law dictating whether people with monocular vision can drive, it is up to each state to determine these regulations.[7] Of course, like everything else, it will take some time and practice to get used to driving with one eye, so you may want to look into specialized driving classes in your area.

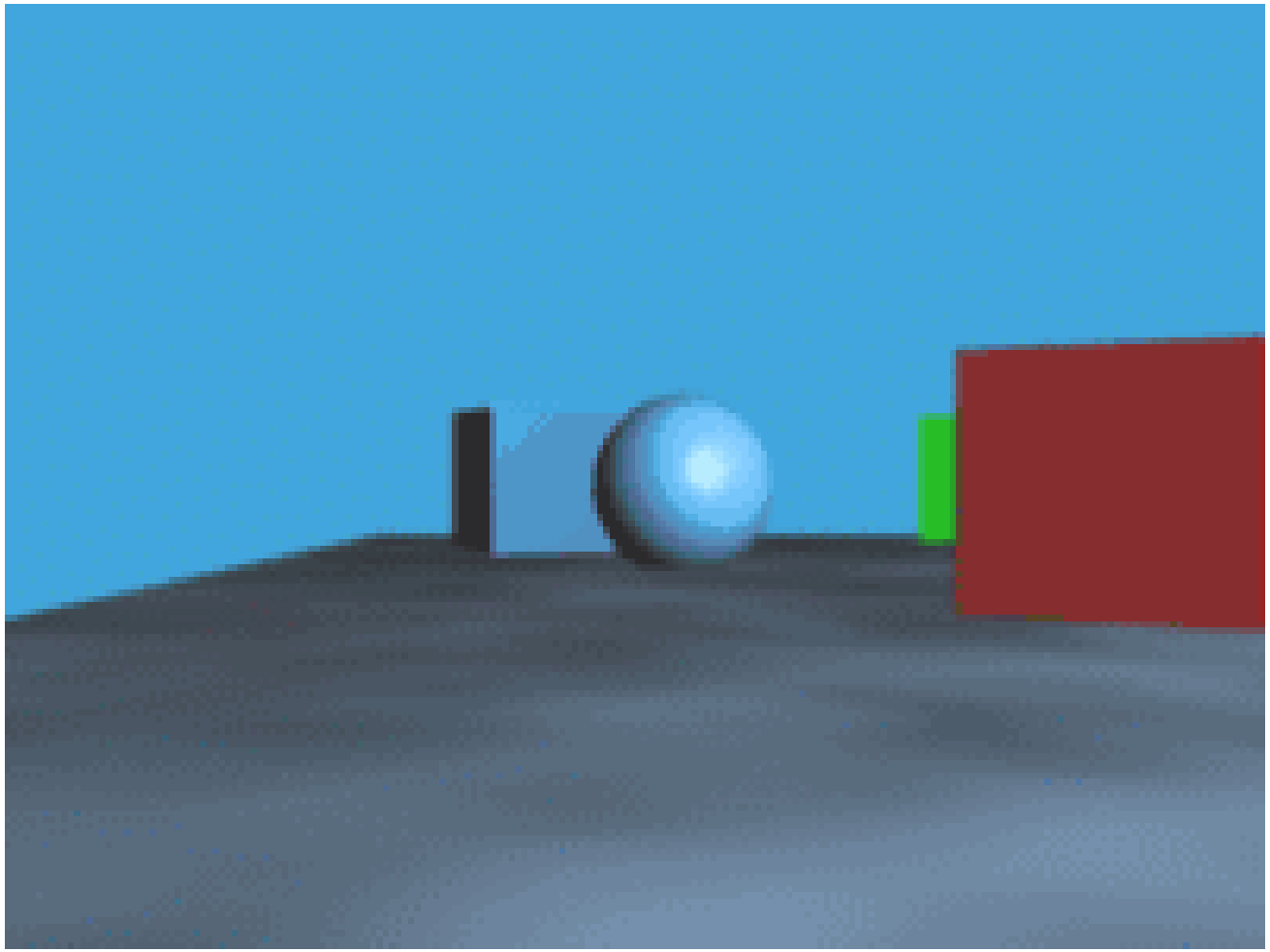
"한 눈 잘 보이면" 시각장애인도 '1종 보통' 면허 딴다

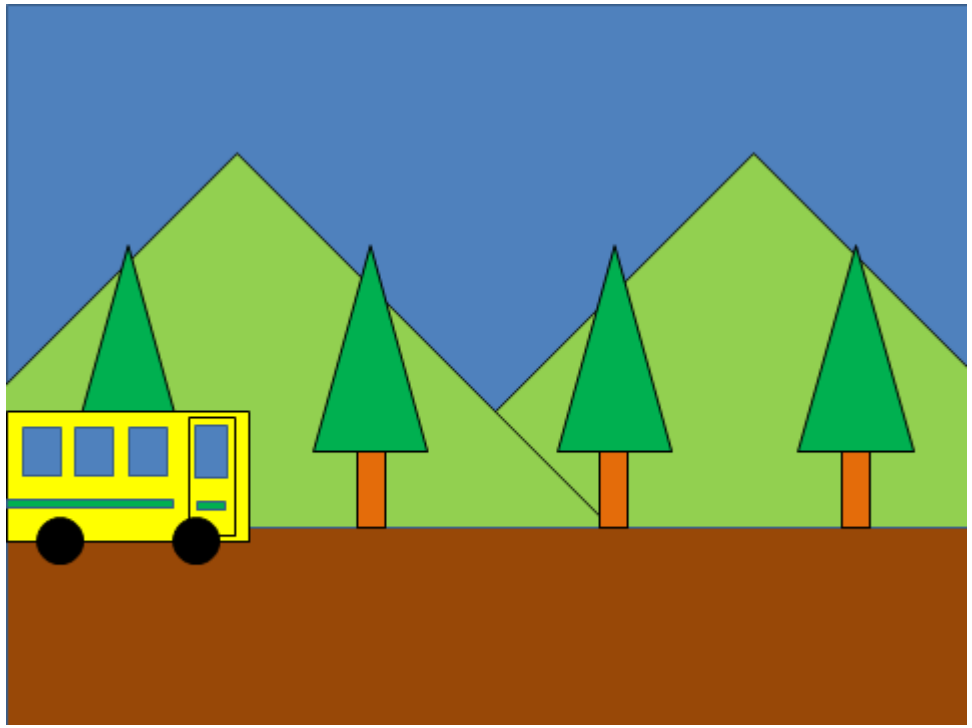
VIEW 19,534 | 2016.04.18 05:20

가 의견 남기기

'시력 0.8·일정 수준 이상 시야' 경찰, 단안 시각장애 면허기준 마련

Monocular Depth Cues (1): Motion





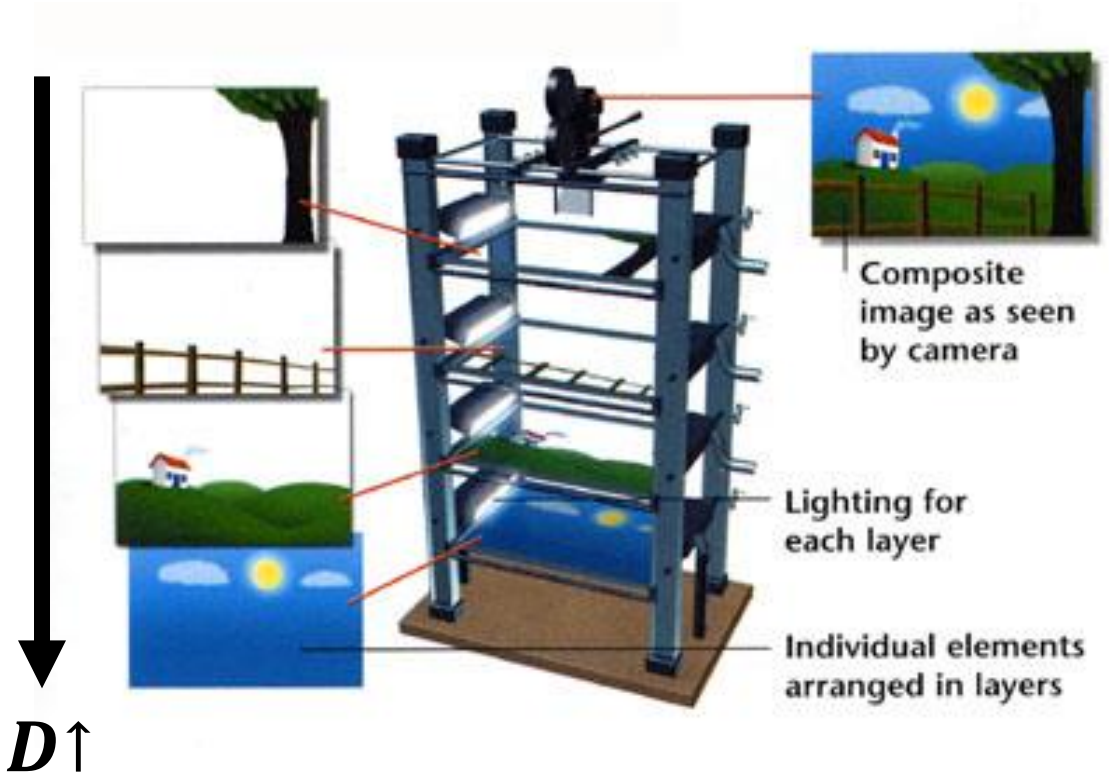
“Motion parallax”

→ Differences in image **motions** between objects at different **depths** (motion ↑ → depth ↓)

[1] Figure credits: Wiki – motion parallax

Motion Parallax in Film Industry

Walt Disney's Multiplane Camera

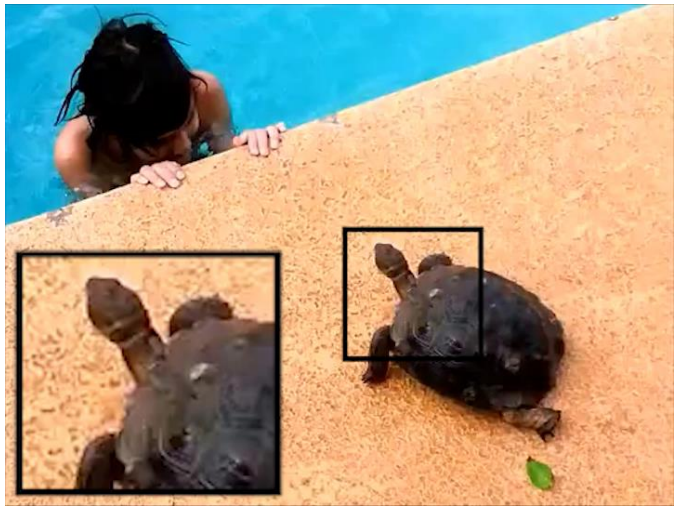


→ Multiplane images allow us to feel the 3D effect !

[1] Walt Disney, "Tricks of Our Trade" episode from Disneyland, February 13, 1957

Motion Parallax by Animals

Strange behaviors of animals…

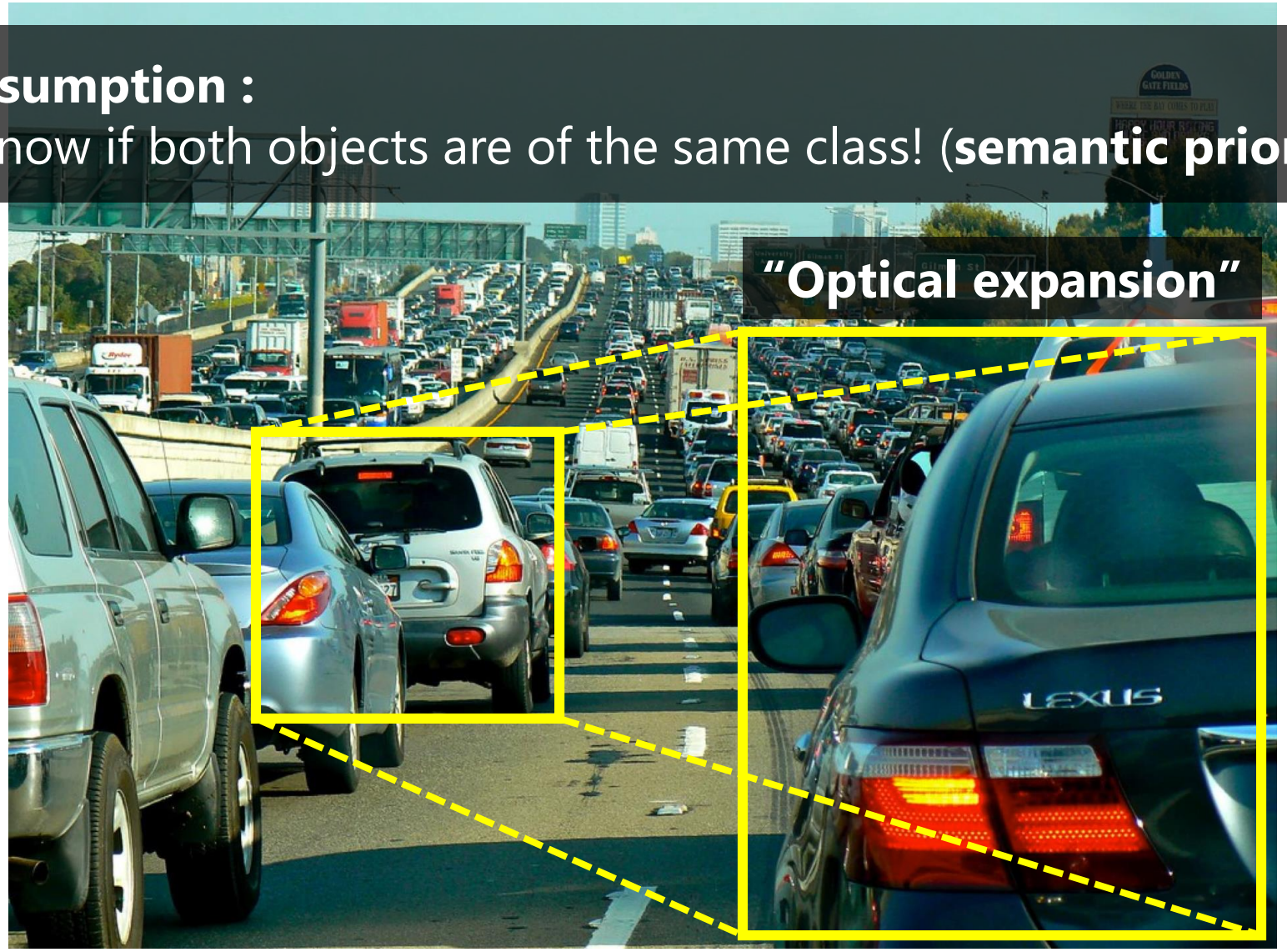


Have a reason… 3D perception!

[1] YouTube, “Getting down with some biggie!”
[2] YouTube, “Super cute budgie bobbing his head”
[2] YouTube, “Headbanging/bobbing Turtle”

Monocular Depth Cues (2): Relative Size

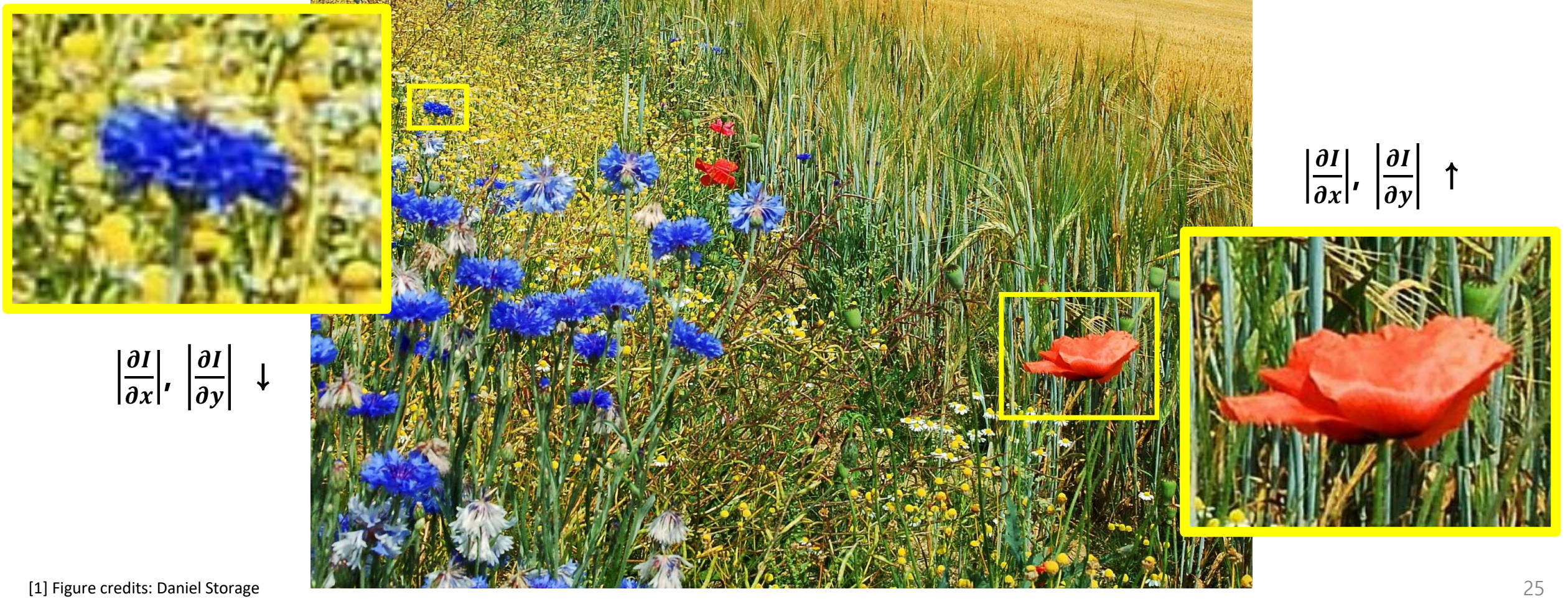
Here is an assumption :
We need to know if both objects are of the same class! (**semantic prior** knowledge)



[1] Figure credits: Daniel Storage

Monocular Depth Cues (3): Texture Gradient

"The farther the object is, the more vague the texture becomes"



$\left|\frac{\partial I}{\partial x}\right|, \left|\frac{\partial I}{\partial y}\right| \downarrow$

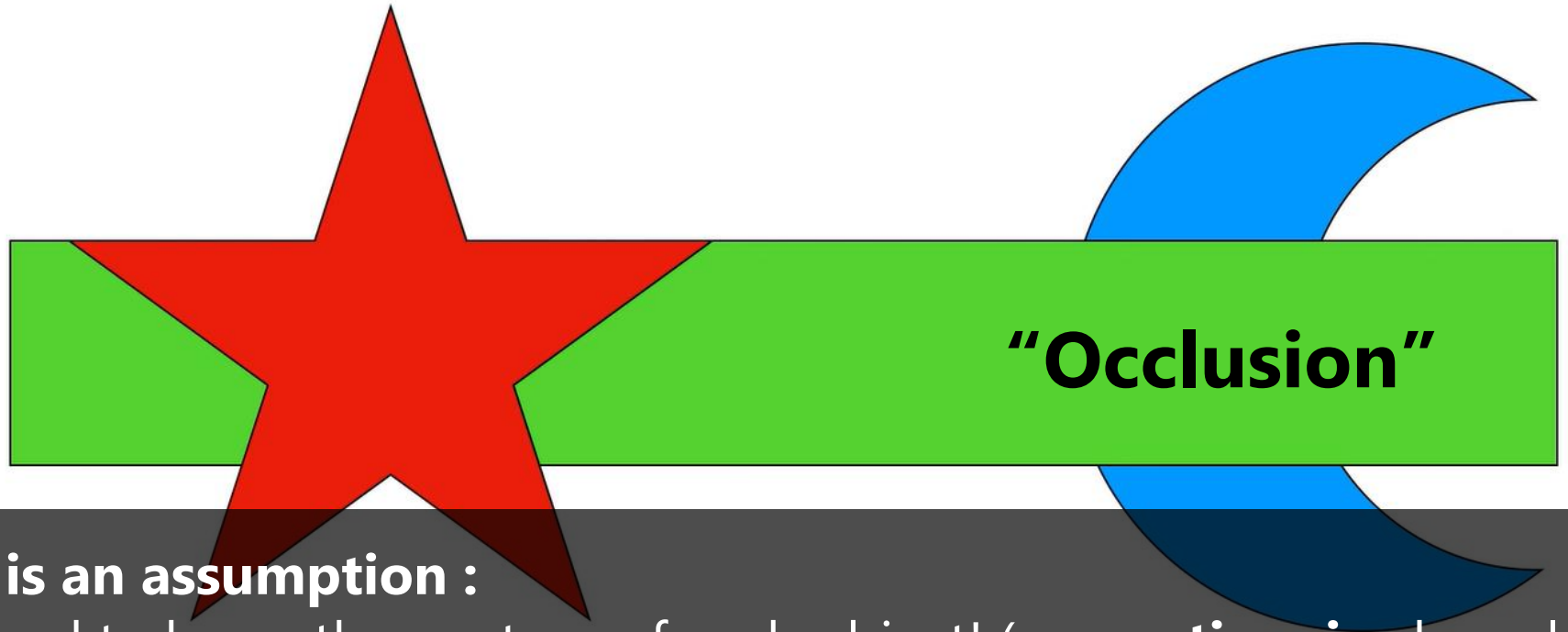
$\left|\frac{\partial I}{\partial x}\right|, \left|\frac{\partial I}{\partial y}\right| \uparrow$

[1] Figure credits: Daniel Storage

Monocular Depth Cues (4): Interposition

Object A is blocking our view of **Object B**

→ **Object A** is closer to us than **Object B**



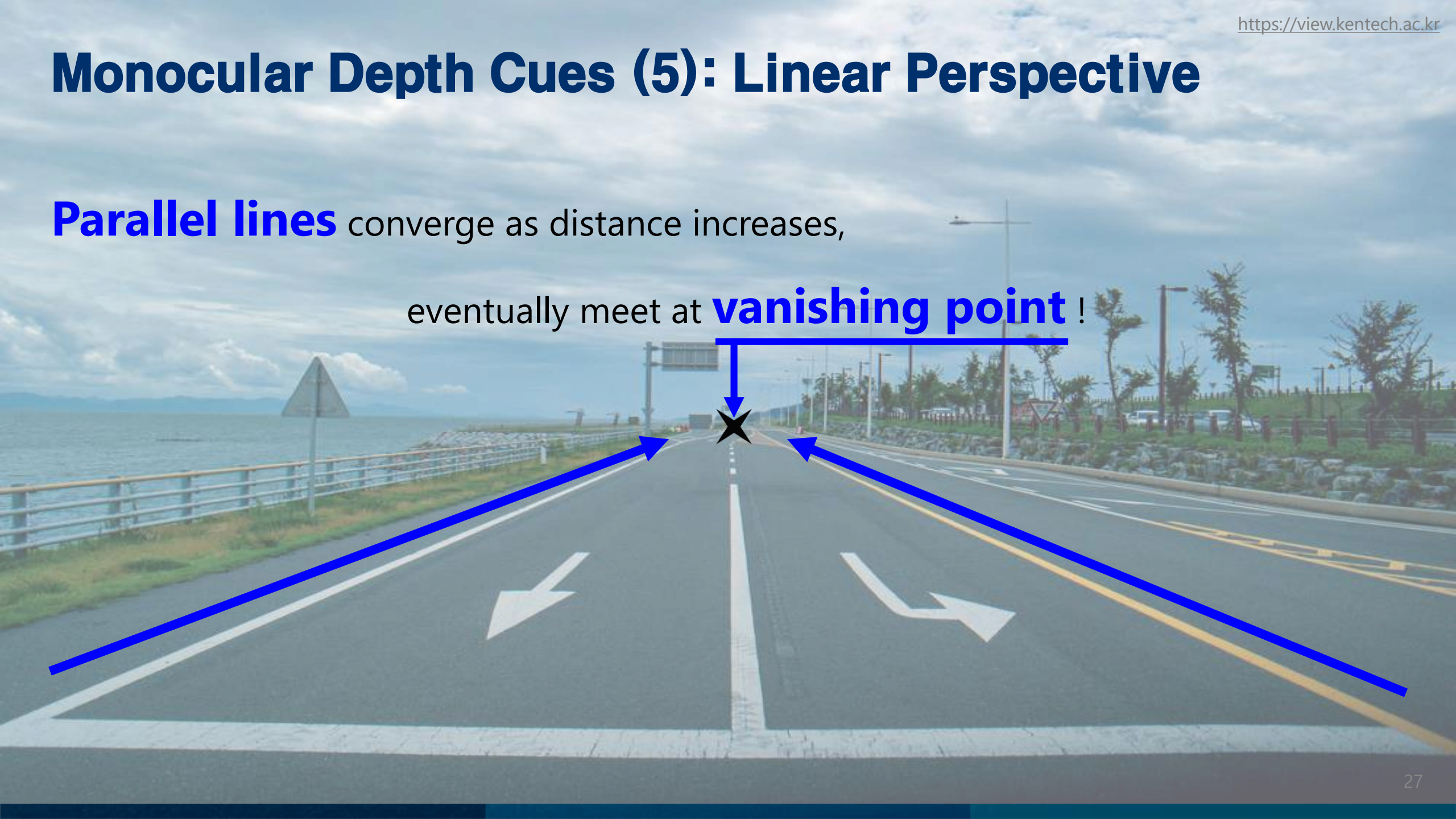
Here is an assumption :
We need to know the contour of each object! (**semantic prior** knowledge)

[1] Figure credits: Daniel Storage

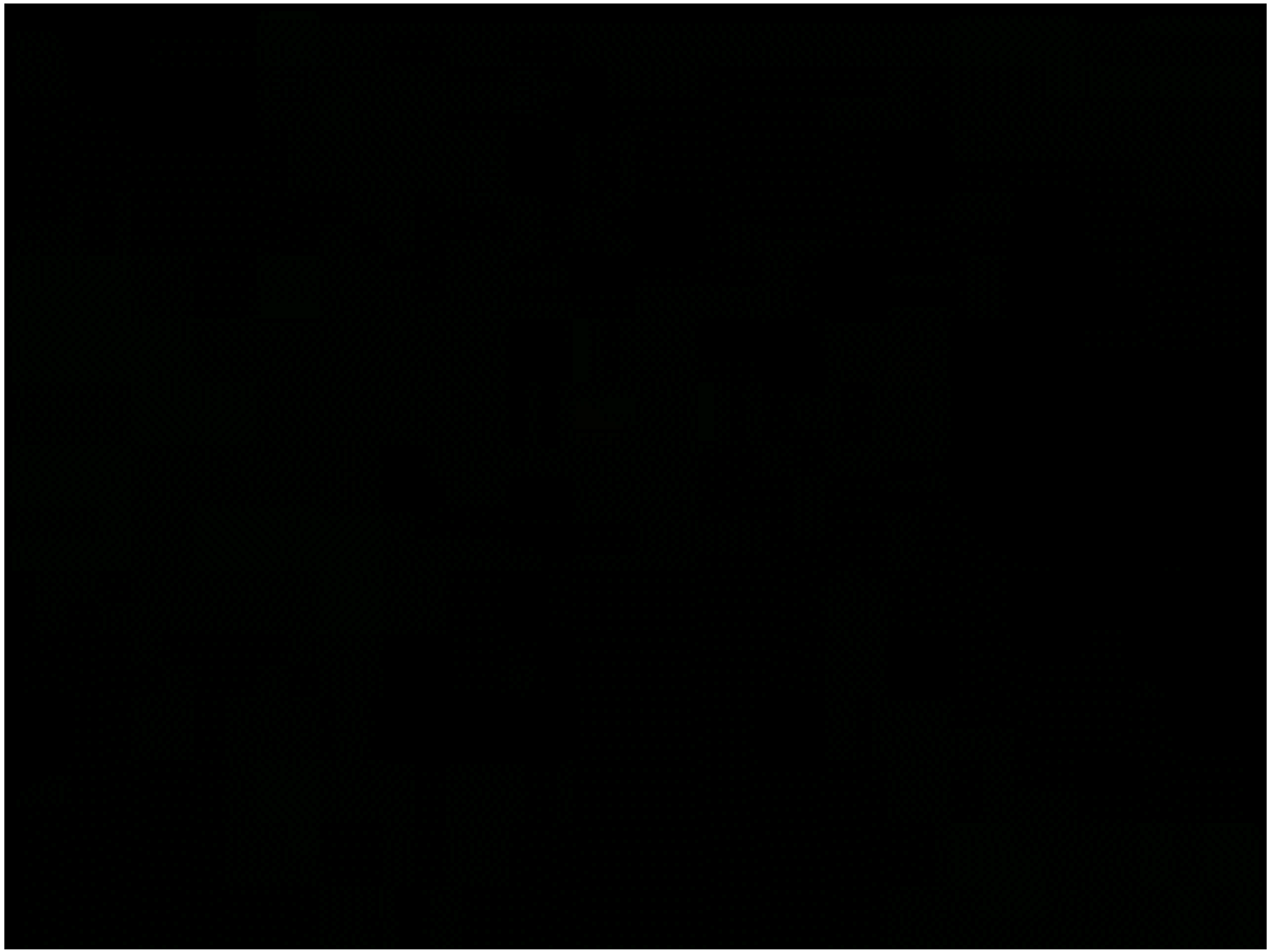
Monocular Depth Cues (5): Linear Perspective

Parallel lines converge as distance increases,

eventually meet at **vanishing point** !



Optical Illusion by False Linear Perspective

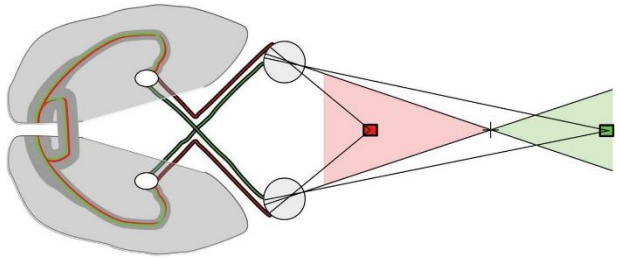


[1] YouTube, “Ames Room (Philip Zimbardo)”

Summary of 3D Perception: from Human to Machine

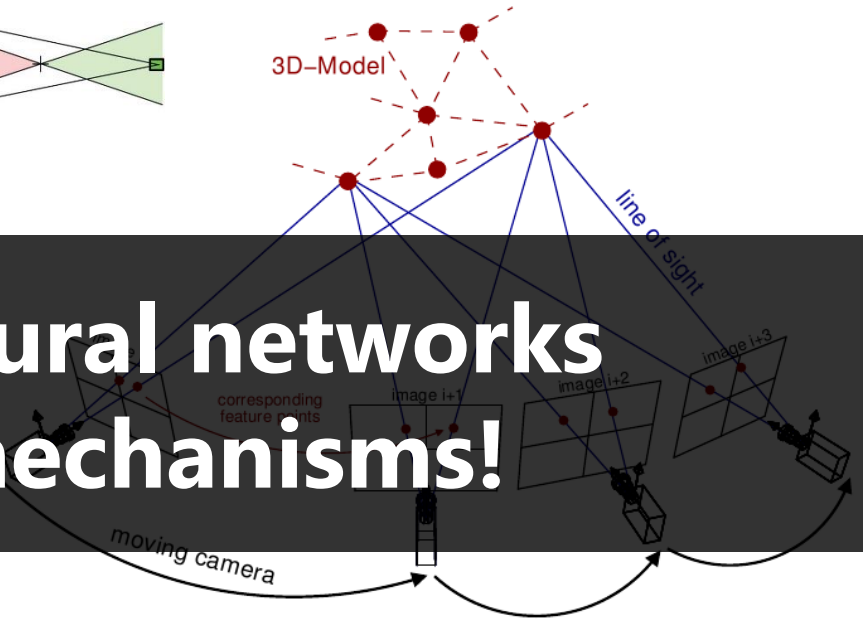
Binocular correspondence

→ "*Stereo matching*" in computer vision



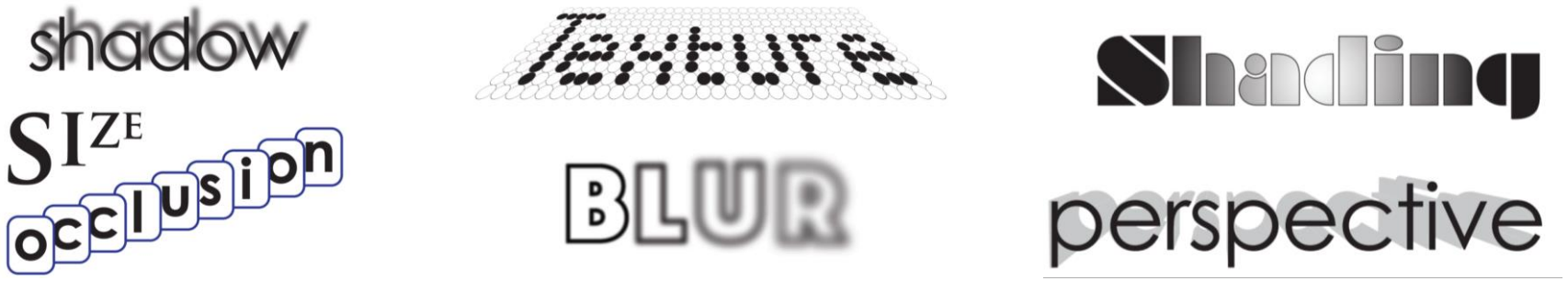
Motion parallax

→ "*Structure-from-Motion (SfM)*" in computer vision



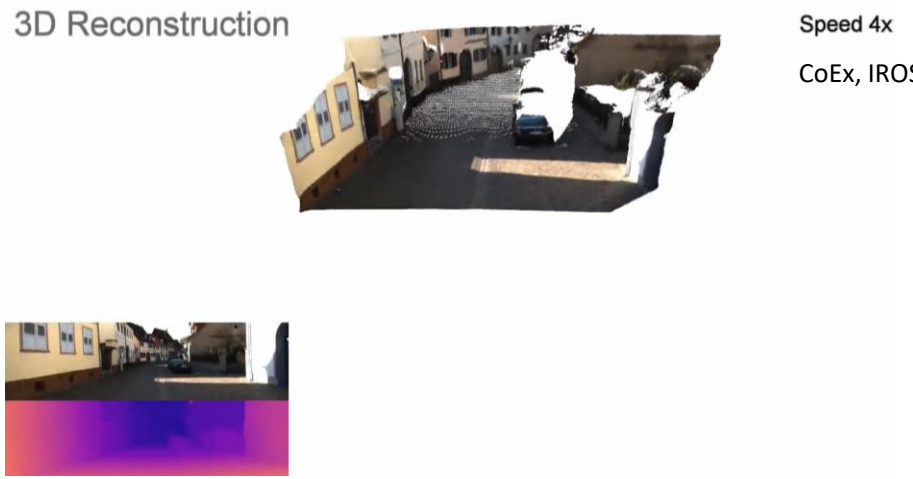
→ Applied to train deep neural networks with different learning mechanisms!

Other monocular cues



[1] A. Welchman et al., "The Human Brain in Depth: How We See in 3D.", Annual Review of Vision Science 2016.

3D Computer Vision Applications



CoEx, IROS'21

Waymo's AI Recreates San Francisco From 2.8 Million Photos! 🚘, 2022

DensePose, CVPR'18

3D human pose estimation

EG3D, 2021

3D face reconstruction


Q&A
KENTECH
Korea Institute of Energy Technology

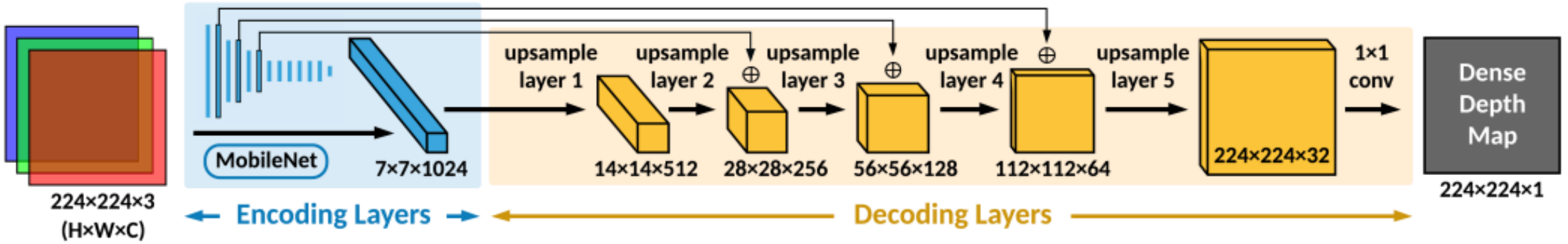
KENTECH
Korea Institute of Energy Technology

(1) (2)

Depth & Pose Estimation on Jetson Nano

Reference – Depth Estimation

Moncular depth estimation (https://github.com/dusty-nv/jetson-inference/blob/master/docs/depthnet.md)



FastDepth, ICRA'19

Reference – Pose Estimation

Human pose estimation (https://github.com/dusty-nv/jetson-inference/blob/master/docs/posenet.md)

→ Locating various body parts (human body + hands)

that form a **skeletal topology** (keypoints + links).

[1] Zhe Cao, et al., "Realtime multi-person 2d pose estimation using part affinity fields." CVPR 2017.
[2] Bin Xiao , et al., "Simple baselines for human pose estimation and tracking." ECCV 2018.

Experiments – Depth & Pose Estimation

*Your basic workspace is here: "**cd ~/jetson-inference/build/aarch64/bin**"

Run mono-depth with image files
Q1. Check the original image file "**images/trail_8.jpg**". Run "**python3 depthnet.py images/trail_8.jpg images/test/output_trail_8.jpg**". Open the result image file, and what is in the file? Please guess the meaning of the different colors in the output figure.

Live depth estimation
Q2. Run "**python3 depthnet.py --network=monodepth-fcn-resnet50 --flip-method=rotate-180**", Please discuss the quality of the results.

Live pose estimation
Q3. Run "**python posenet.py --flip-method=rotate-180 csi://0**", and "**python posenet.py --network=resnet18-hand --flip-method=rotate-180 csi://0**". Try different networks and discuss how each output is different. What does each node (keypoint) and link mean?
You can check the models in the below link:
https://github.com/dusty-nv/jetson-inference/blob/master/docs/posenet.md#pre-trained-pose-estimation-models

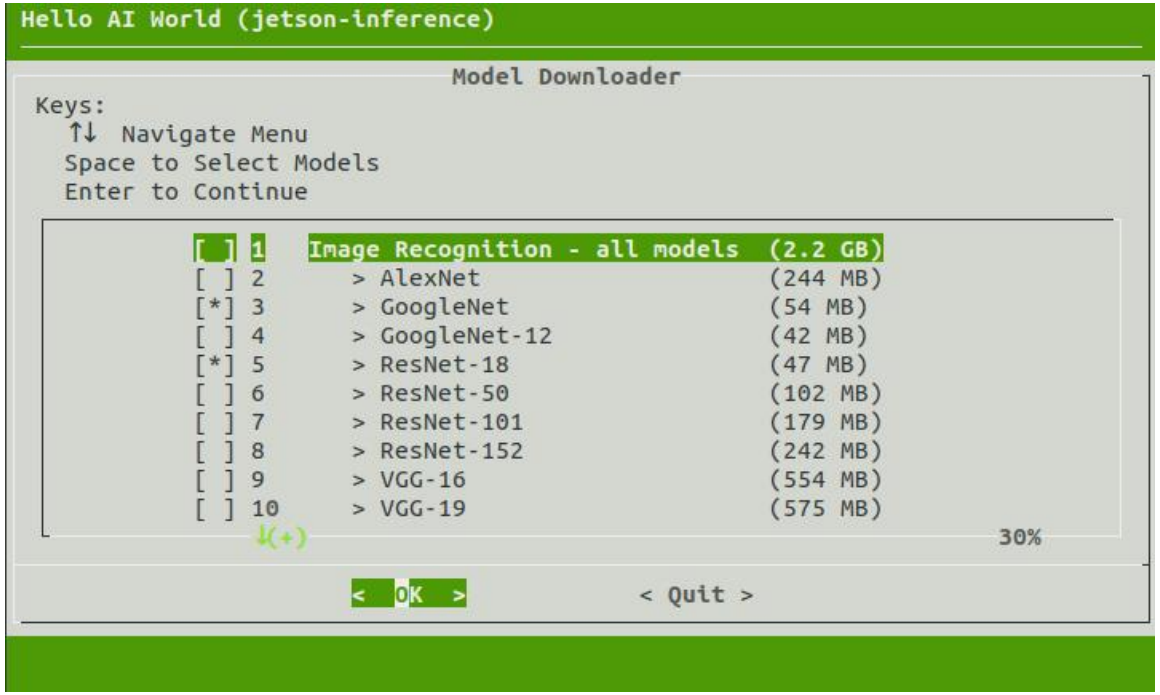
Experiments – Depth & Pose Estimation

How to try other models?
Check your installed model,
"**cd ~/jetson-inference/build/aarch64/bin/networks**"
"**ls**"
you can see the model folder installed.

Download models,
"**cd ~/jetson-inference/tools/**"
"**./download-models.sh**"

After download,
"**cd ~/jetson-inference/build/aarch64/bin**"

→ Submit all experimental results (classification, detection, segmentation, depth, pose) in **one** report.
→ The deadline is postponed until **Nov. 14th (Mon)**.



Tesla's Autopilot is Now on Humanoid Robot

Semantic + 3D Visual Perception

In the **Tesla AI Day 2022** (about a month ago),

https://youtu.be/ODSJsviD_SU