University of Illinois at Urbana-Champaign
Dept. of Electrical and Computer Engineering

ECE 101: Computing Technologies and the Internet of Things

Computer Vision

Vision is a Powerful Tool

How will modern computer vision be used?

To understand where we might go,
let's look to the past…

Powerful Vision has Appealed for Millenia

Many human mythologies
- include beings associated with
- **enormous powers of vision**,

Sometimes through use of many eyes, as with Argos.

Or through extremely powerful sight, as with Heimdall.

What if You Could See Everything, All the Time?

What would you do

- if you had **eyes everywhere**,
- **all the time**?

Obviously, that's too much information to process oneself.

But now imagine that it is filtered, so you **see only what is interesting**.

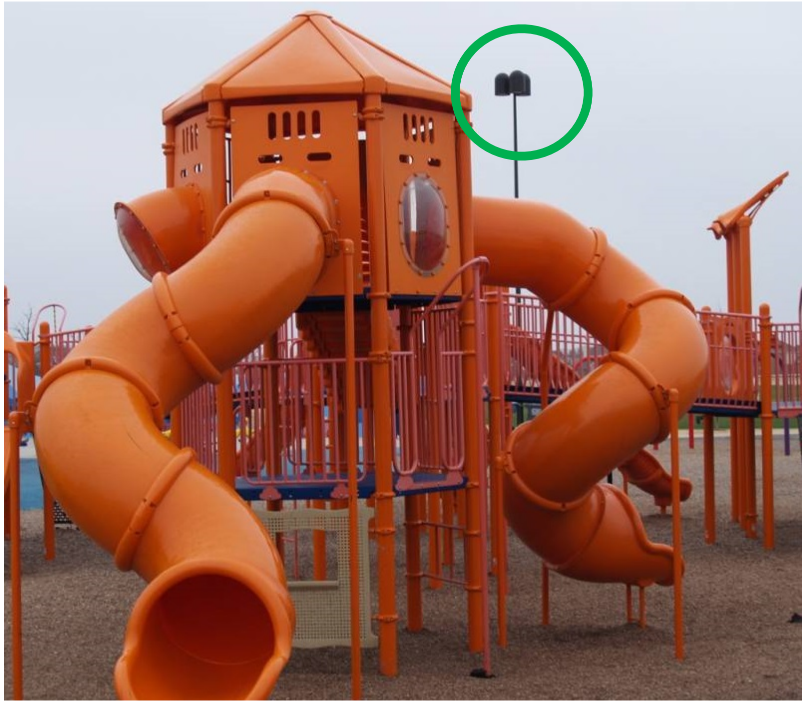
What will you do?

Parenting? Let us Lend an Eye!

Kid-tracking cameras

- **ensure** that your **child** is **safe** from harm and
- **notifies you** instantly **of** any **unusual activity** or behavior!

You can finally relax!



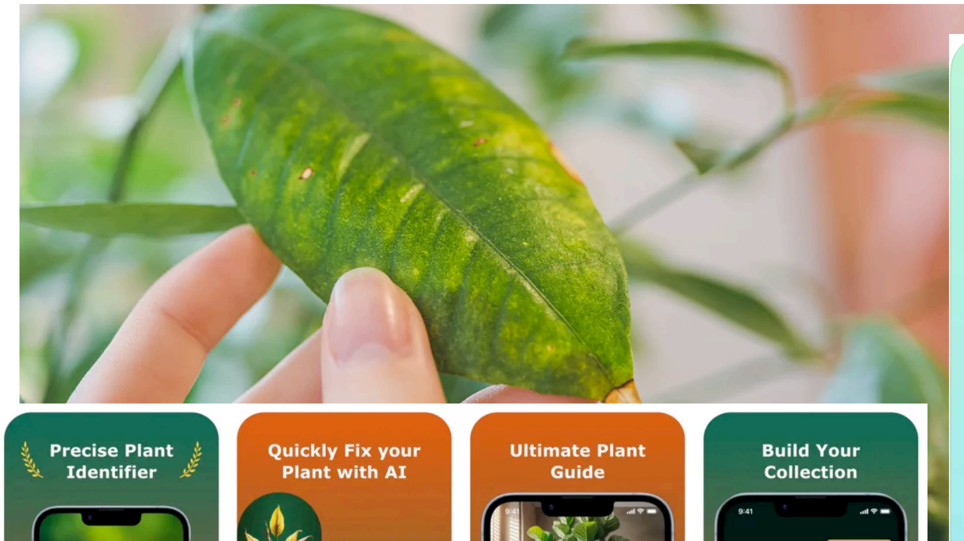
Surveillance and Security: For your Peace of Mind

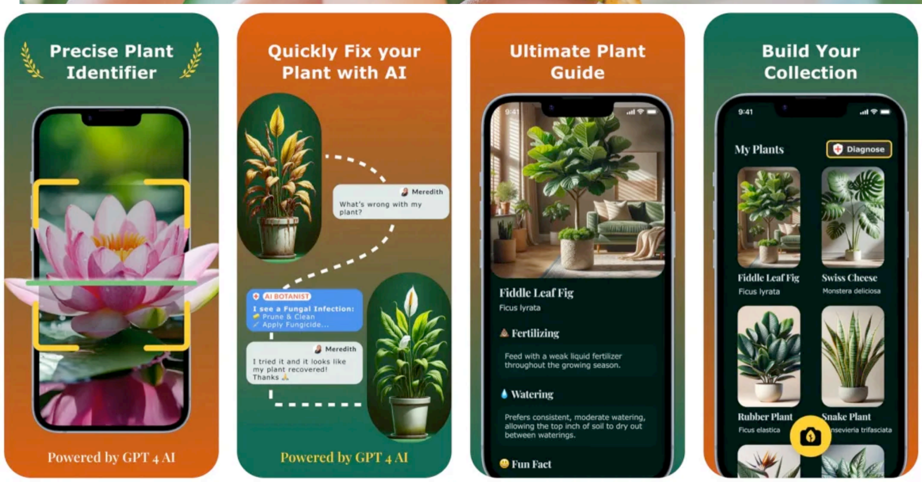
No Checkout Lines: Just Walk Out

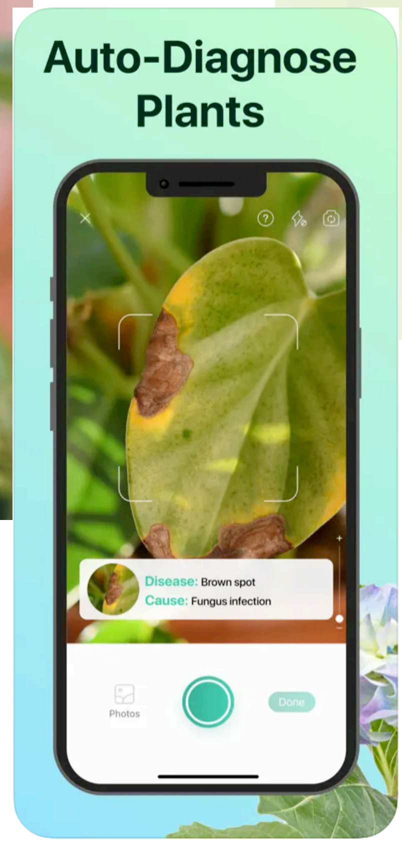
https://www.aboutamazon.com/news/retail/amazon-just-walk-out-dash-cart-grocery-shopping-checkout-stores

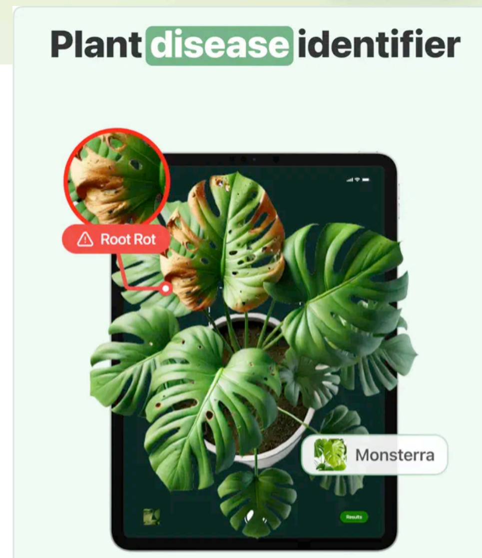
- Computer vision, specifically item and action recognition
- Sophisticated low-cost sensors for weight and vision,
- Deep Learning ML models

Taking Care of your Plants





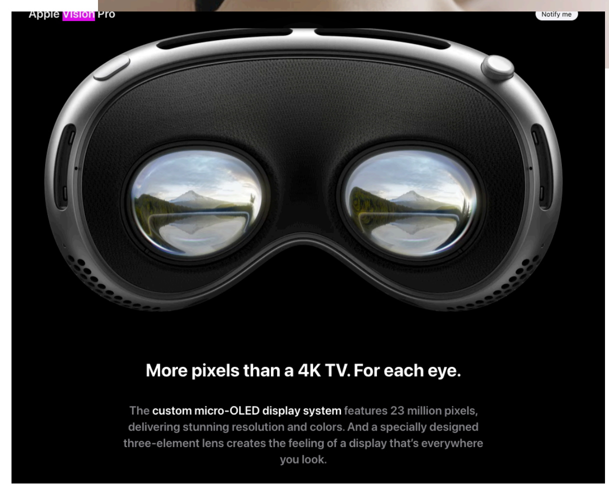




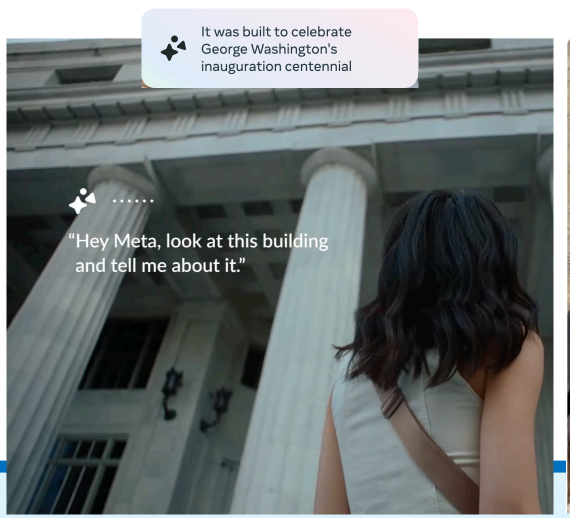
Virtual Vision: This is Your Life!

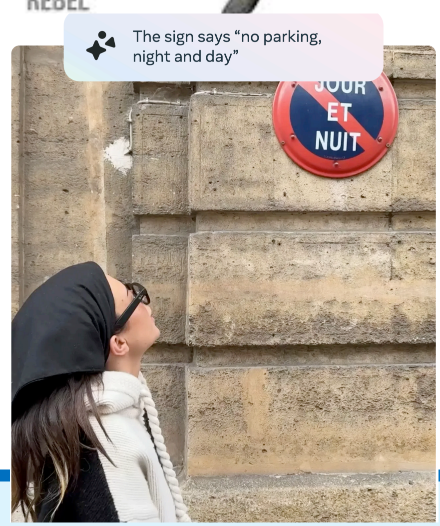
The **LiDAR Scanner** and **TrueDepth camera** work together to create a fused 3D map of your surroundings, enabling Apple Vision Pro to render digital content accurately in your space.

EyeSight reveals your eyes and lets those nearby know when you're using apps or fully immersed in an experience. When someone approaches,
Apple Vision Pro **simultaneously lets you see the person and reveals your eyes to them**.



Smart Glasses





Robots Also Need Vision!

Of course, robots will also need vision.

Robots that look and act like humans are not many years away.

Processing visual signals
- will be an important part in enabling such robots
- to act and communicate with humans.

Whether they look like or unlike us seems to be a matter of choice.

What if You Could See Everything, All the Time?

What will you do?

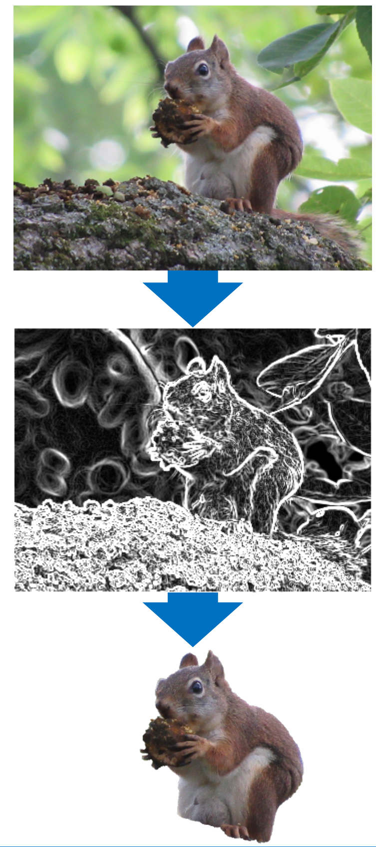
Answer on Canvas— Explain with the help of Sense, Computer Communicate, Actuate loop

Sense, Compute, Communicate, Actuate

What about computation?

It might require **multiple steps**:

1. **combine sensor data** over time
2. **identify and classify** types of objects/entities
3. **understand the physics** of the things identified



Computation: Systems Need to Learn and Predict Behavior

Think about young animals learning to hunt—they need to study the motion of their prey.

If we want systems **to act "intelligently,"**

- the systems **need to learn the differences**
- between humans, dogs, squirrels, cars, trucks, and so forth.

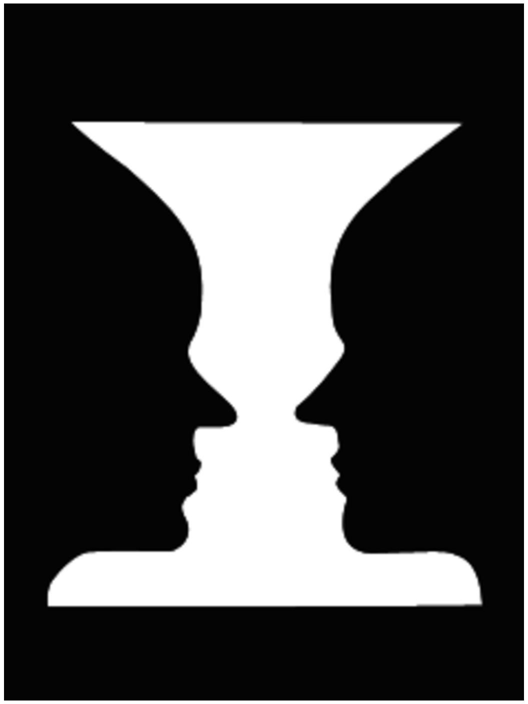
Puzzles Play with Our Brain's Visual Processing

What do you see in the image on the right?

Two faces?

Or a vase?

Most people can see both if they try.

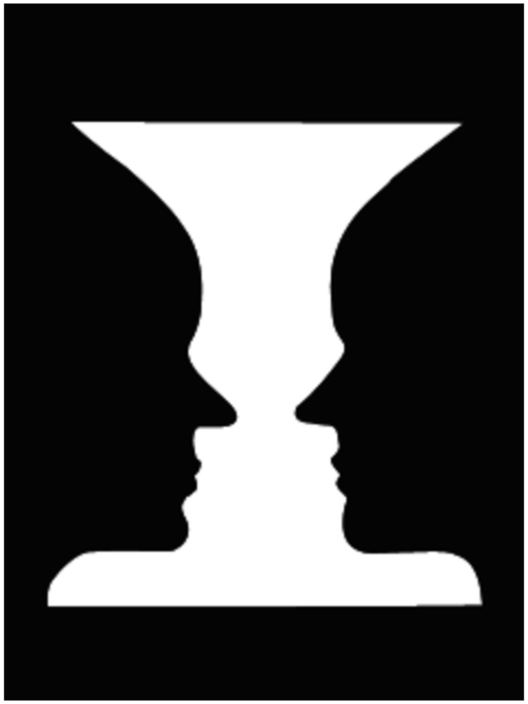


Puzzle Missing Clues Found in the Real World

Why can we see both?

The **visual clues** our brains use are **missing or undermined**:

- consistent **coloring** (forced here)
- **depth** perception (no depth in image)
- **shading** (removed here)
- **structure** (none visible)
- **motion** (faces move like faces; vases don't move)

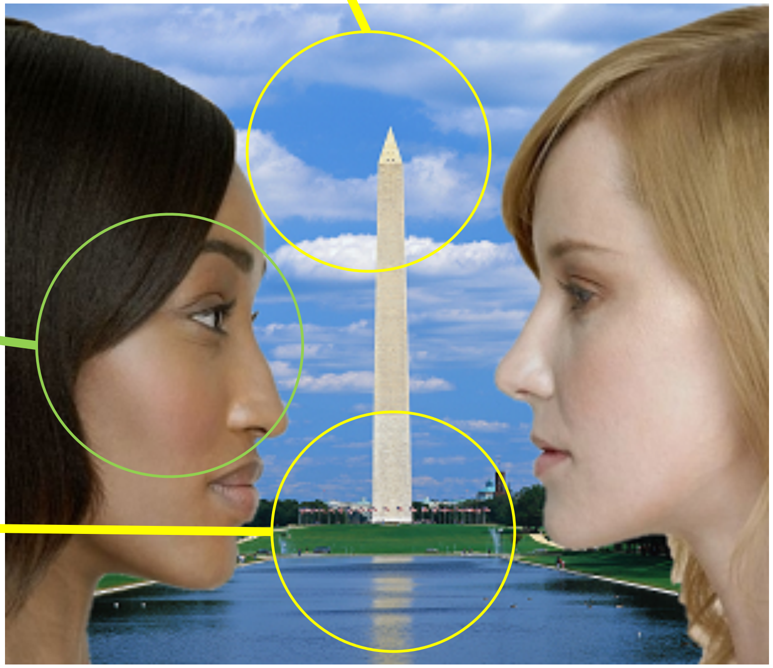


A More Realistic Image has Clues

Do you see a vase now?

(You can, but you really have to use your imagination!)

Really colorful vase?



Color, shading, and structure match our idea of a face.

Shading and structure not right for a vase shape.

Tools and Techniques Used by Our Brains

What goes on in our brain to enable us to tell the difference?

Some of everything we talked about:

- **edge detection** by light/color,
- **segmentation** into areas of common color,
- use of depth perception and shading to **estimate 3D structure**, and
- **motion tracking** to estimate body dynamics.

Can We Be Fooled by Statues?

Body dynamics

A realistically colored and shaped
- **dog statue can trick us**
- until we decide that
- it should have moved but did not.

Look carefully and you may notice
- the “fur” doesn’t look like fur;
- your brain is inferring material properties from lighting.

Modern Sensors Make these Tasks Easier

Accurate interpretation of vision **requires lots of information**.

Modern technology does make the task easier.

Rather than depending

- on purely **passive sensors** (cameras),
- we **make use of active sensors**, such as lidar and radar.

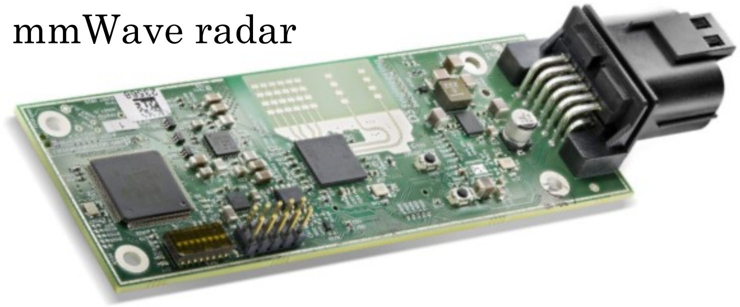
Sense, Compute, Communicate, Actuate (Vision)

Consider monitoring a street intersection.

What is being sensed?

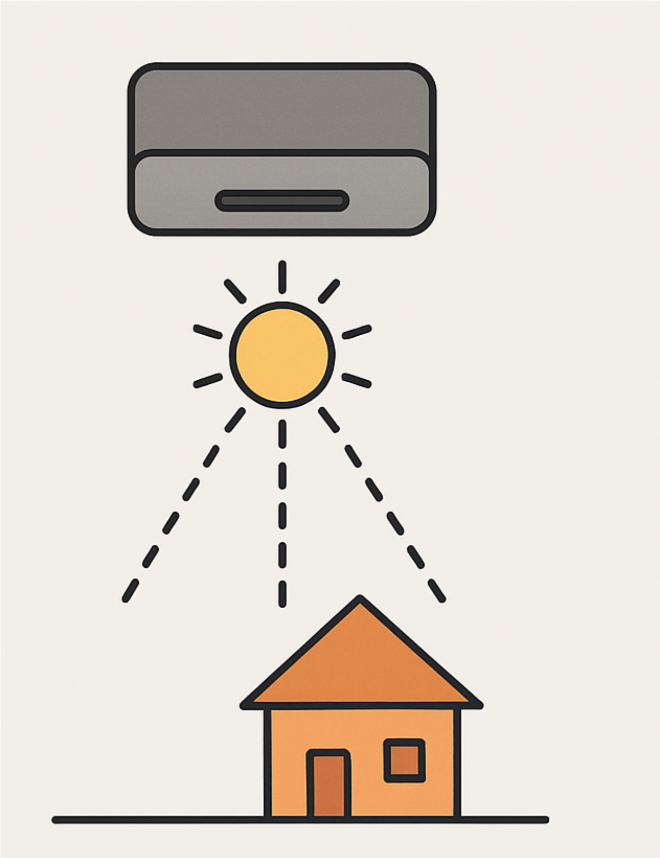
- **"Visual" information from**
  - **both passive**
  - **and active sensors.**

mmWave radar



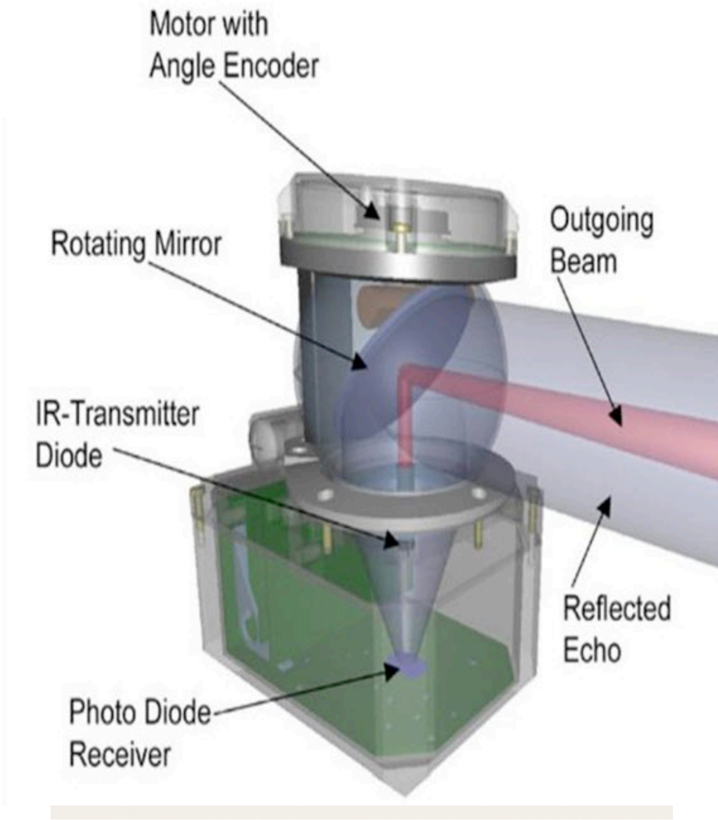
Passive Sensors

- Do not **emit** any energy
- **Rely on ambient light or natural radiation**
- Capture data by detecting light (visible, infrared) or other electromagnetic waves already present.
- **Examples**:
  - Standard RGB cameras
  - Infrared cameras (thermal imaging)
  - Satellite imaging using sunlight
- **Advantages**:
  - Simpler and consume less power.
  - Ideal for daylight or well-lit environments.

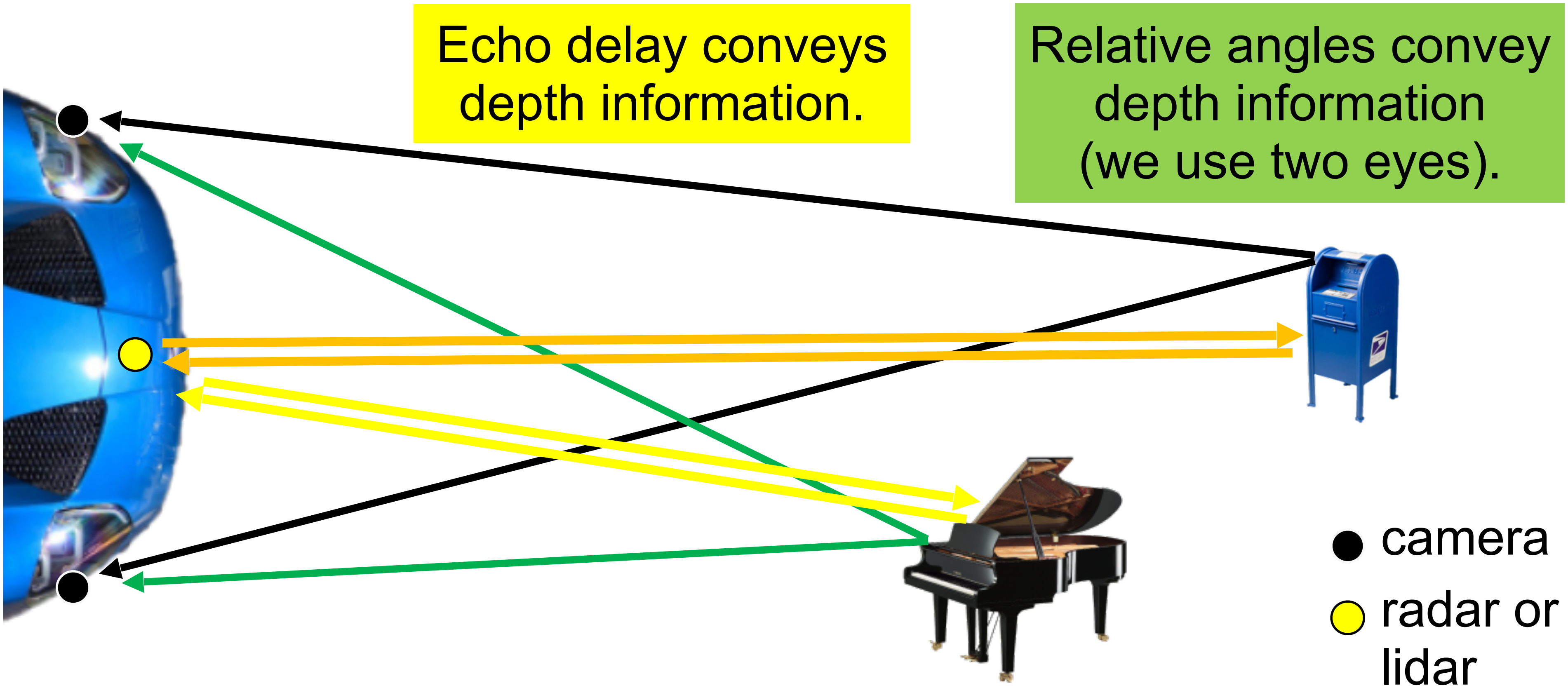


Active Sensors

- **Emit energy** (like light, sound, or radio waves) and measure the reflection or response.
- Can operate in darkness or controlled conditions
- Examples:
  - LiDAR (Light Detection and Ranging)
  - Radar
  - Structured light sensors (used in depth cameras like Kinect)
- **Advantages**:
  - Useful for depth sensing,
  - 3D mapping, and
  - low-light or night-time applications.



Depth Can be Gauged Passively or Actively



Can a Tesla be Fooled by a Screen

Depth Perception Adds to Active Sensor Information

Why use depth perception?

Active sensors rely on specific frequencies,
- which **may not reflect** from all materials,
- **and may require** correct **shape**.

For example,
- some radar (NOT mmWave) reflects only from metals
- stealth aircraft use surfaces angled to avoid reflection

Feature Registration Enables Fusion of Sensor Data

Modern systems use both

- **active** sensors generate "**point clouds**" (with depth labels), **and**
- **passive sensors** to generate **images**.

Specific points on objects

- must be **identified and equated** (e.g. between image and point cloud)
- in order to **combine the sensor information**.
- This is called feature registration

Example:

- Which pixels/points correspond to the mailbox and which belong to the cat?
- Which pixels/points correspond to the left front leg of the mailbox?

Sensor Fusion

Combining data from

different types of sensors

to create a more accurate, robust, and comprehensive understanding of the environment.

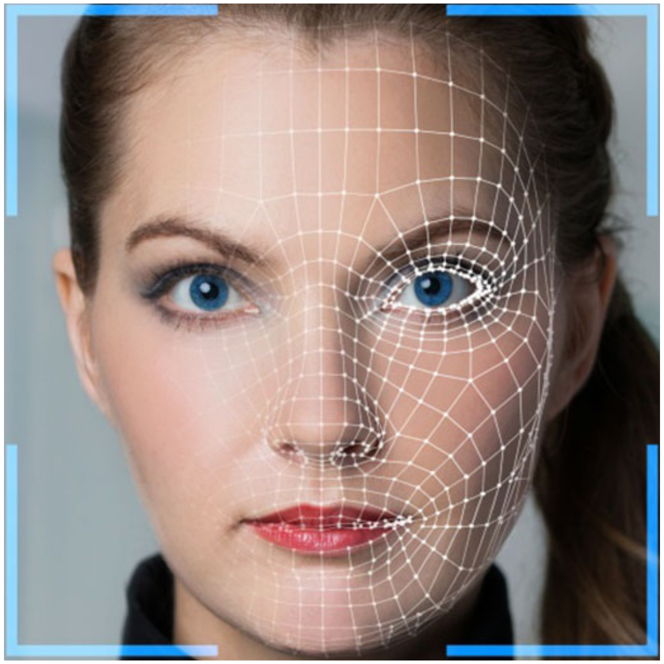
Feature Identification Enables Recognition of Individuals

Feature extraction / registration

- also **used for recognizing** things
- and **instances of things**.

Example: given a human face,

- extract a set of points:
- nostrils, eyebrows, lips, and so forth.
- relative positions of features
- allow matching of the individual face.



Must Understand **Motion** to Integrate Data Over Time

All **sensors** may **suffer from occlusion**:

- an object is partially or completely
- **behind something else**.

One **"instant" of data**

- allows us to **build**
- a **partial 3D map** of the surroundings.

To combine information over time, we **must**

- identify objects and **associate** them
- **with types of movement** (mailbox: doesn't move!).

Computation Example: Edge Detection

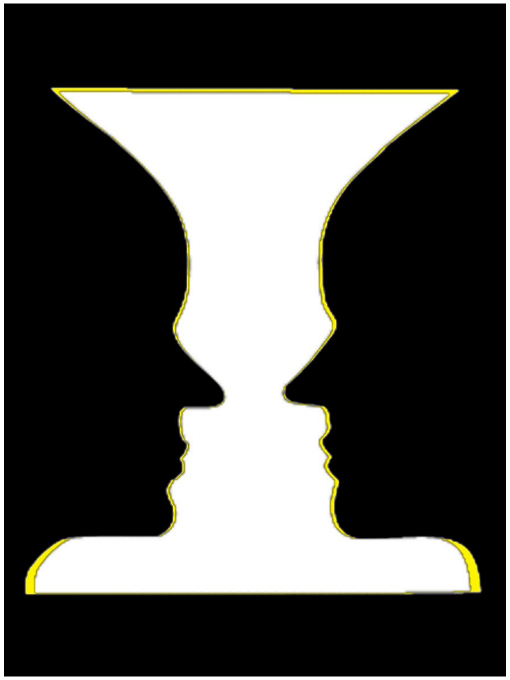
Let's look more closely at one of
the first steps: **edge detection—f**ind boundaries
between segments in images.

In humans,

- the first level of neurons in the visual cortex
- performs similar processing,
- **finding the boundaries between different parts of an image**.

For example, in the image to the right,

- the yellow region is the "edge"
- between the vase and the rest of the image.

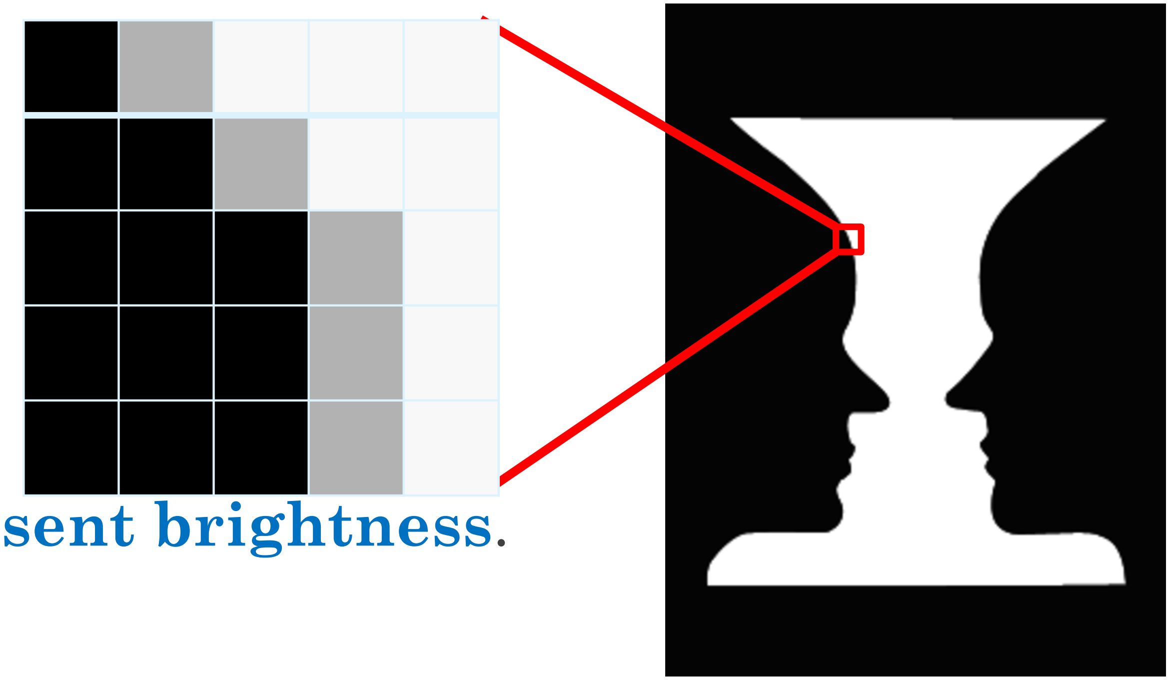


Monochrome Pixel Values Represent Image Brightness

Let's first zoom in to a small image section.

0	1	2	2	2
0	0	1	2	2
0	0	0	1	2
0	0	0	1	2
0	0	0	1	2

=



The pixel **values represent brightness**.

Filter Captures Notion of a Vertical Edge

Let’s use this **3×3 filter** of values.*

For each pixel,
- we center the 3×3 filter
- over the pixel,
- multiply each overlapping pixel
- with a filter value, and
- sum all of the results.

The sum indicates whether a vertical line appears at that pixel.

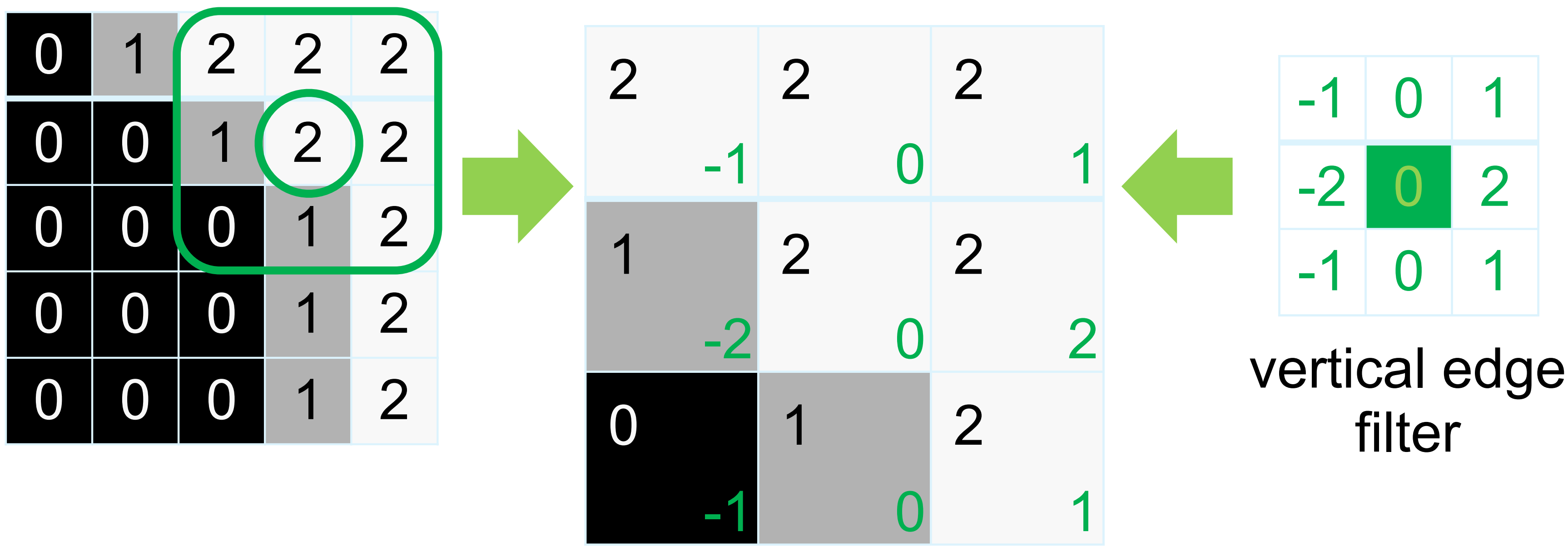
*The filter was designed by I. Sobel and G. Feldman in 1968.

-1	0	1
-2	0	2
-1	0	1

vertical edge filter

Example of Computing One Pixel's Vertical Edge Value

Choose a pixel, align, …

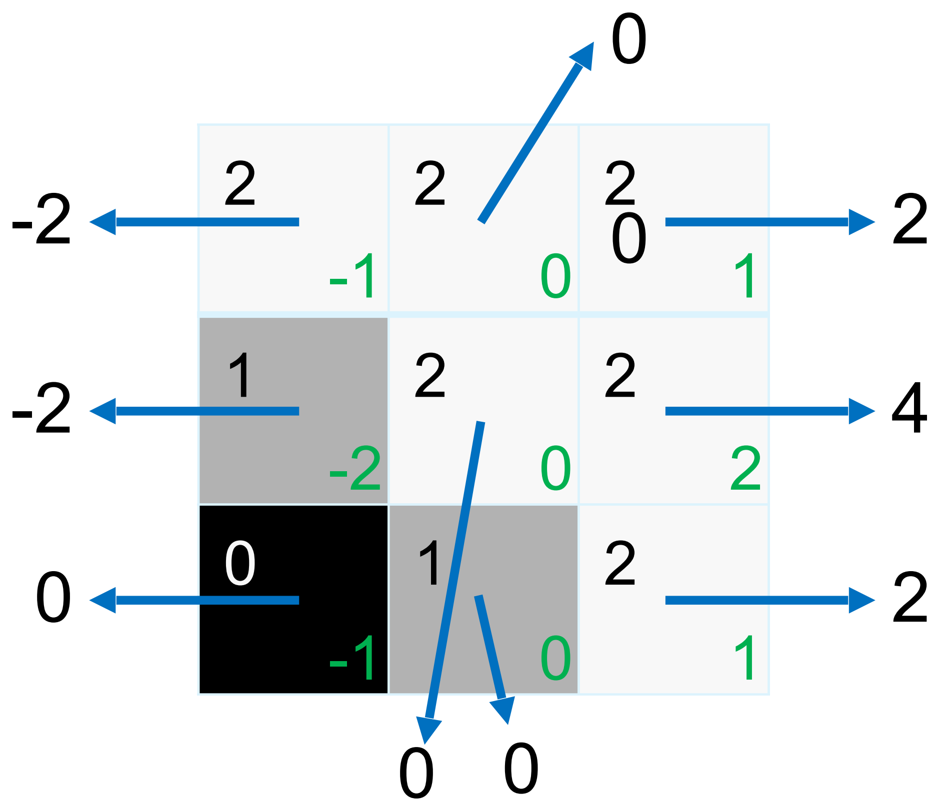


Continuing the Computation of One Pixel's Value

… multiply …

… and sum …

Answer is 4!



The Computed Values Highlight Vertical Lines

Do the same for every pixel (called **convolution** with the filter).

0	1	2	2	2
0	0	1	2	2
0	0	0	1	2
0	0	0	1	2
0	0	0	1	2

3	7	5	1	0
1	4	6	4	1
0	1	5	7	3
0	0	4	8	4
0	0	4	8	4

strongest vertical line (other parts of the line are not vertical)

(I duplicated the boundary cells for computation.)

With Measure for Horizontal Lines, Find Line Direction

A second 3×3 filter detects horizontal lines.

By combining the two results,
- we **obtain line strength**
- **and** line **direction** at each pixel.

-1	-2	-1
0	0	0
1	2	1

horizontal edge filter

-1	0	1
-2	0	2
-1	0	1

vertical edge filter

Line Directions (Perpendicular) Identify Segment Bounds

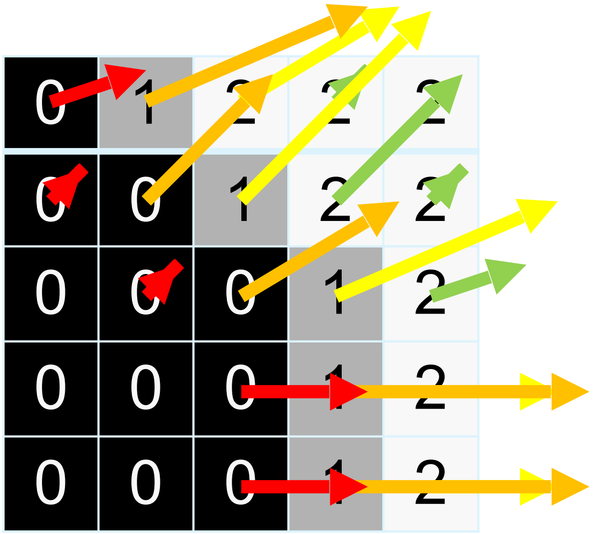
The vectors drawn to the right

- illustrate the result of the filters
- (colors are just for clarity).

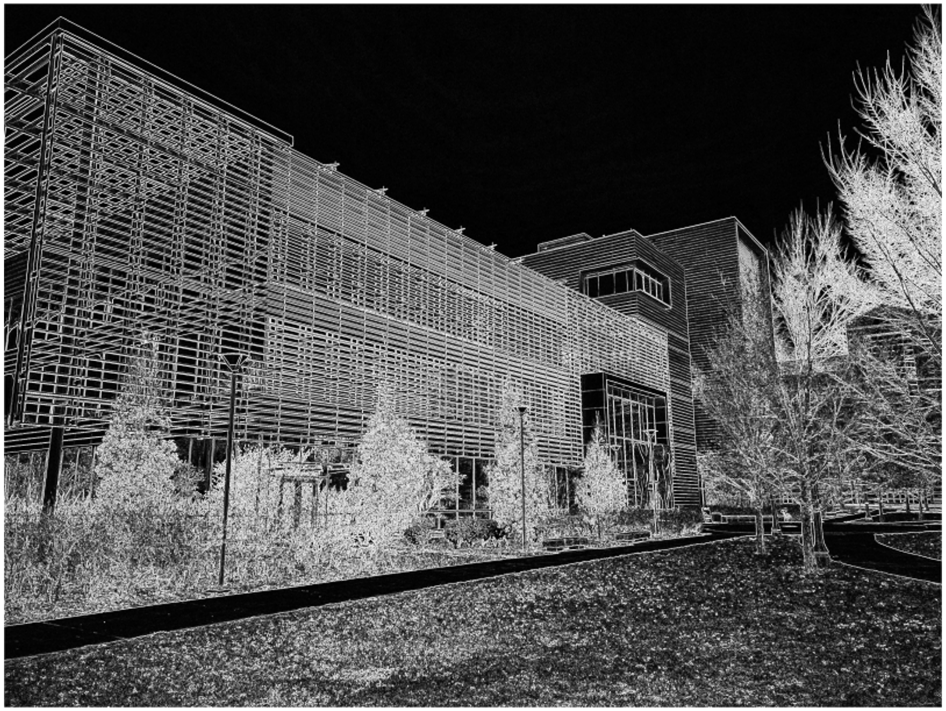
By **finding continuous sequences**

- of adjacent vectors
- **with similar direction**,
- we find curves that split the image.

The **curves bound segments** of the image, **which correspond to objects**.



Example: Magnitude of Edge Vectors of ECEB



Example: Magnitude of Edge Vectors of CSL

Boundaries Define Segments Corresponding to Objects

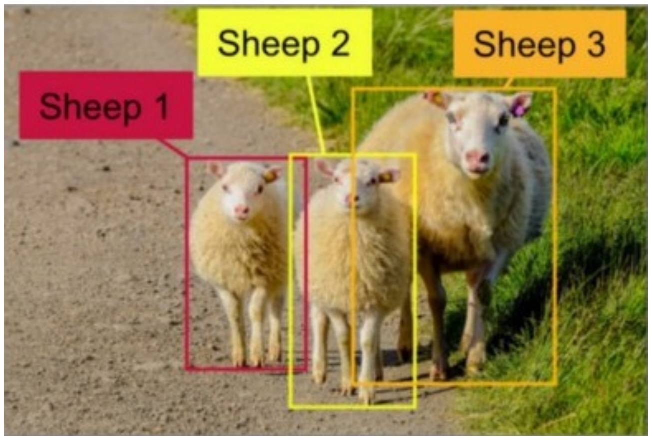
Using the boundaries,
- we **split** an **image into segments**.
- In our simple image, there are only two.

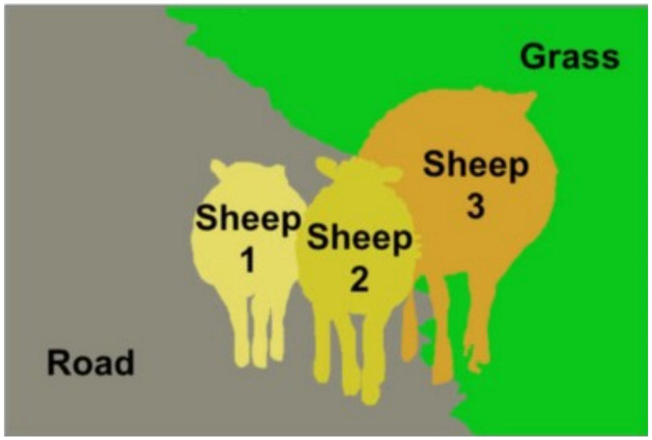
Segments
- can be **matched against object models**
- to find appropriate **labels**.

ML Filters (Convolutions!) Can Identify Regions

Or we can use ML to train convolutional filters

- that **mark regions according to object type**,
- then look for the regions within the segmented image.





Fewer Object Types Means Easier Identification

As with speech,
- a **smaller "vocabulary" of objects**
- **makes identification easier**.

For example, one might omit
- models of lions and elephants
- from many systems.

And almost no systems need to be able
- to spot a rampaging Tyrannosaurus Rex,
- despite the fact that most humans could
- (or so they think).

Some Objects Take Time to Identify Correctly

“At first it seemed a little speck,
And then it seemed a mist;
It moved and moved, and took at last,
A certain shape, I wist.”
—Samuel Taylor Coleridge,
The Rime of the Ancient Mariner

Sometimes, it **takes time and observations**

- of the motion of an object
- in order **to identify** it.

ML and Traditional Models May Mix

Towards this end,
- when correct identification matters,
- the ML-based part of the system
- is likely to be backed
- by a more traditional set of object models.

As with our brains,
- these **models** can be **used in both directions**:
- feed **shapes and movement in**to the model **to classify an object**, **or**,
- **given** an object's **classification, predict** the kind of **motions** that it can make.

“Visual” Data Also Fused with Other Types

“Visual” data are likely to be **integrated with other types** of information.

Let’s say that my car is following another vehicle, for example.

How do we know when the vehicle in front slows down?

1. **Brake lights**
2. **WiFi notification** from vehicle (in future)
3. “Visual” **speed/depth change** detection

Most Sensors Must be Treated as Unreliable

1. Brake lights
2. WiFi notification from vehicle (in future)
3. "Visual" speed/depth change detection

The **first two may fail**:

- slowing may occur without braking, and
- brake lights can go out or be obscured.
- WiFi notifications may not arrive, or may not be sent from older cars.

Trust What You See?

1. Brake lights
2. Notification from vehicle (in future)
3. "Visual" speed/depth change detection

Only change detection is reliable

- (and even there, we need to have multiple sensors—ours can also fail),
- but it is **also delayed**—takes time to notice.

Safety-Critical Systems Require Safety Envelopes

In practice, systems require a "**safety envelope**":

- make sure that the car
- has **enough time to react safely**
- **to changes** in the environment.

Safe Decision Making Relies on Layers of Models and Data

Information on slowdown

- must be fed into **models for motion planning**,
- which in turn rely on accurate **models of vehicle dynamics**,
- which rely on sensing the **vehicle's load**.

Average car weight is 4,000 pounds, with a cargo load of about 1,000 pounds.

That's a 25% change in mass, assuming humans obey rules.

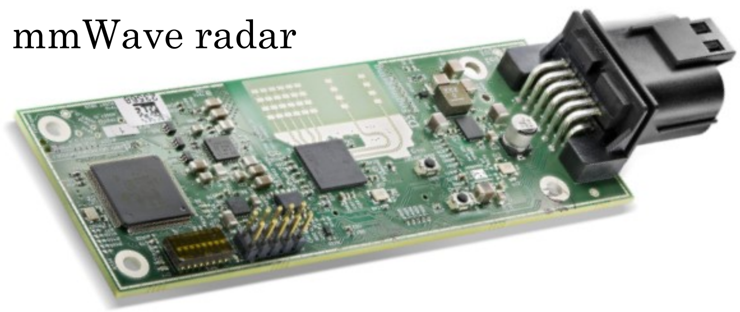
The Last Phases: Communication and Actuation

Let's close the loop by returning to the cycle: sense, compute, **communicate** AND **actuate**.

Once a vision-capable system has interpreted what it sees, **it can respond**.

Each system will be different:
- **drive** the vehicle **safely**,
- **authenticate humans** and **report unusual activity** or
- **deliver messages** when human returns home.

mmWave radar



Terminology You Should Know from These Slides

- active vs. passive sensors
- edge detection
- segmentation
- segments (of an image)
- depth perception
- motion tracking
- point cloud vs. image
- feature registration
- occlusion
- filter
- convolution
- safety envelope

Concepts You Should Know from These Slides

- human use of visual clues to identify objects
- why depth perception is useful together with active sensors
- use of feature extraction to recognize individuals
- why understanding movement is useful to "intelligence"
- integration of information over time and across sensors
- role of safety envelope in ensuring safety