GE 461 Introduction to Data Science

Spring 2023

Applications: Computer Vision

Hamdi Dibeklioğlu

Computer Vision

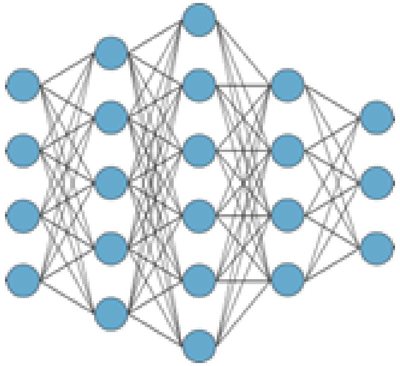
According to Hollywood

Computer Vision

Scene

Sensor

Interpreter



Interpretation

Mountain
Lake
Trees
Sky

Human Vision

Scene

Sensor

Interpreter

Interpretation

Mountain
Lake
Trees
Sky

Computer Vision

Computer Vision

Computer vision tries to get computers to extract information from images

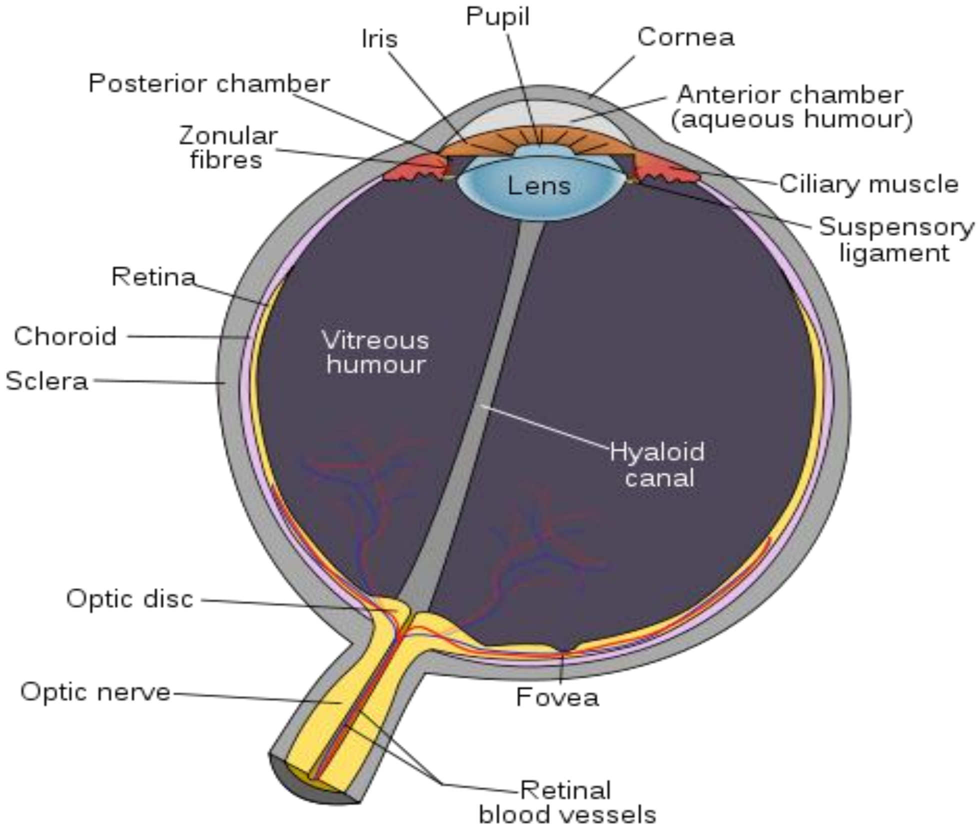
Human Vision

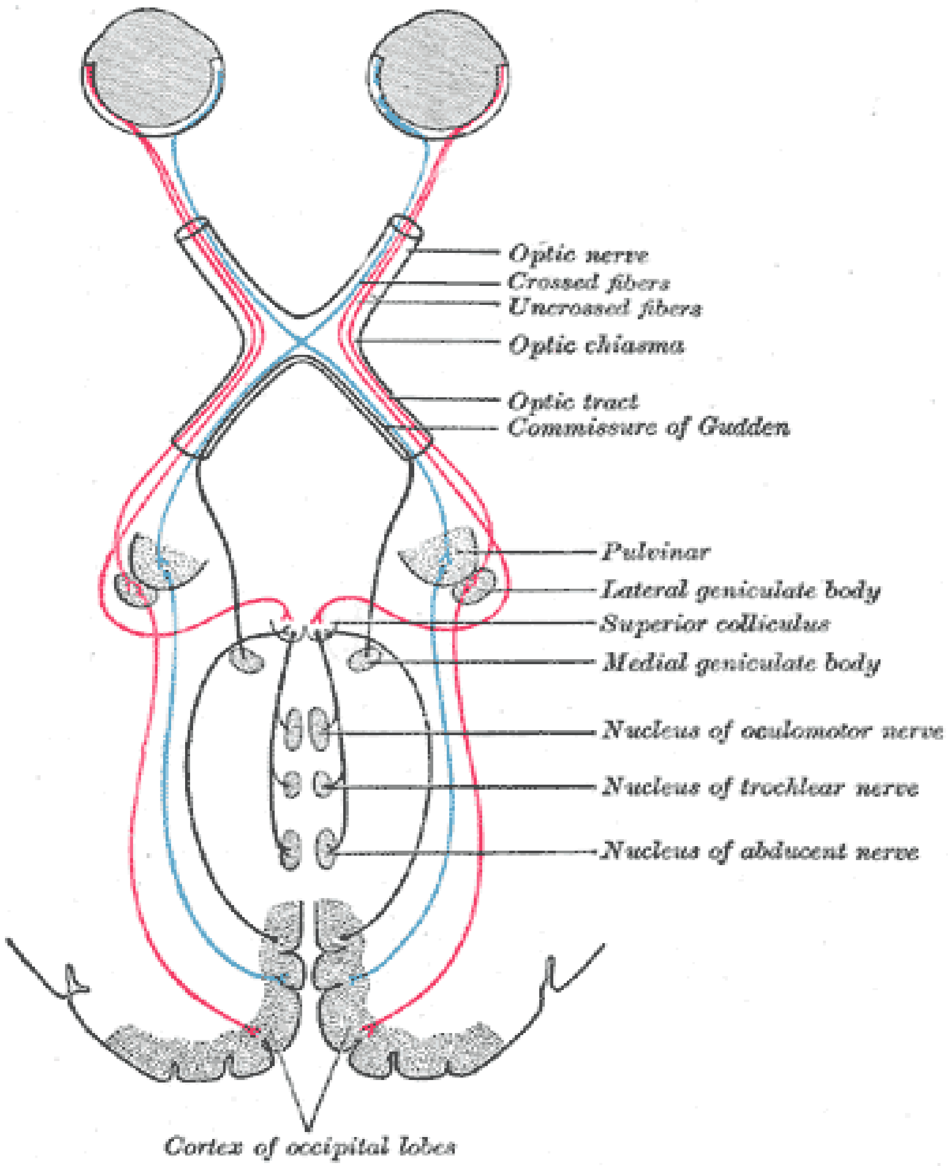
- Can you name some things that influence what we see?

Human Vision

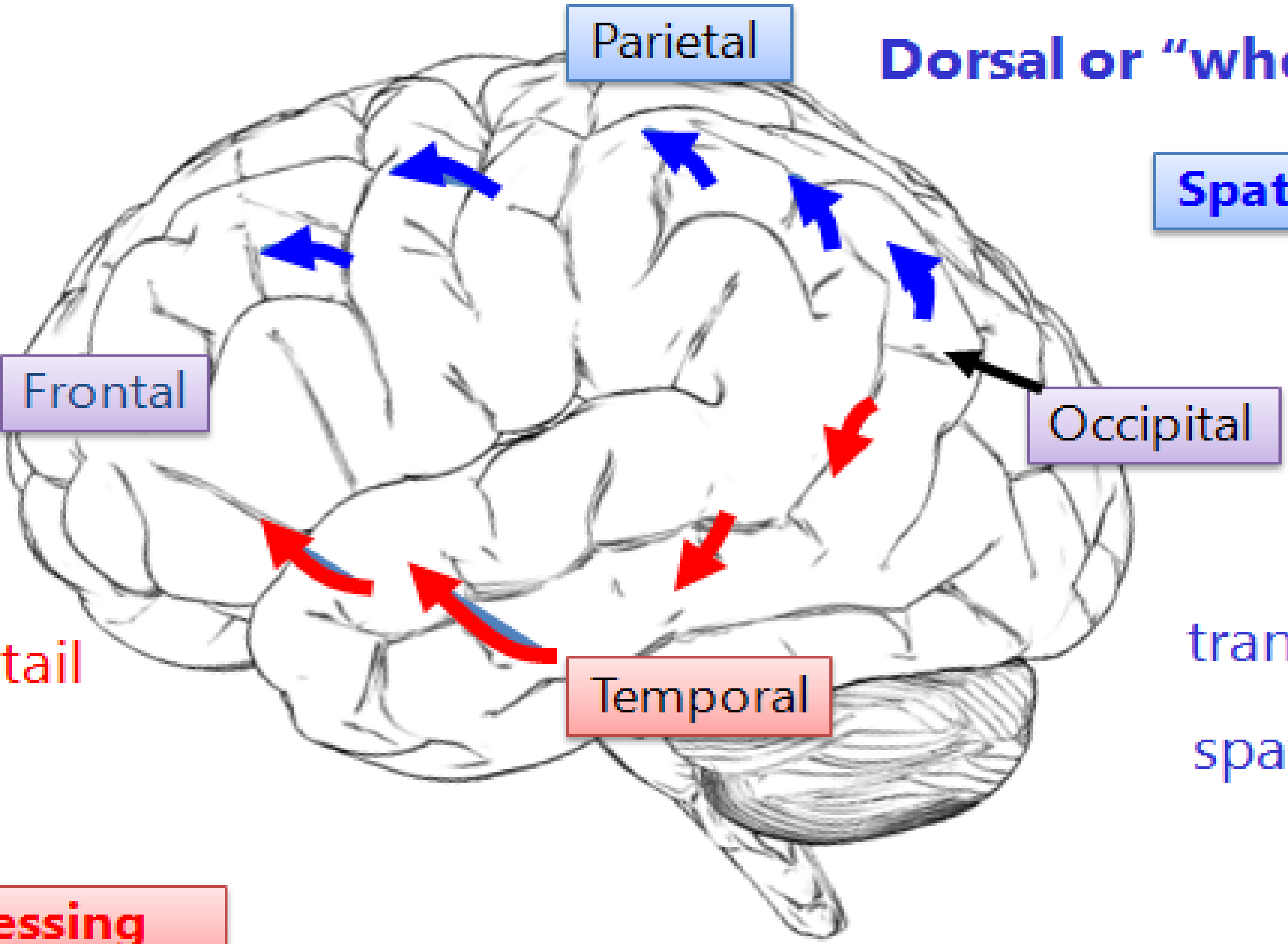
- Can you name some things that influence what we see?
  - What objects are where in the world (and how they are deformed)
  - Lighting conditions of the environment
  - Position and orientation of the eyes (viewpoint)
  - Your own brain!!!

Human Vision





Human Vision



Dorsal or "where" stream

Spatial processing

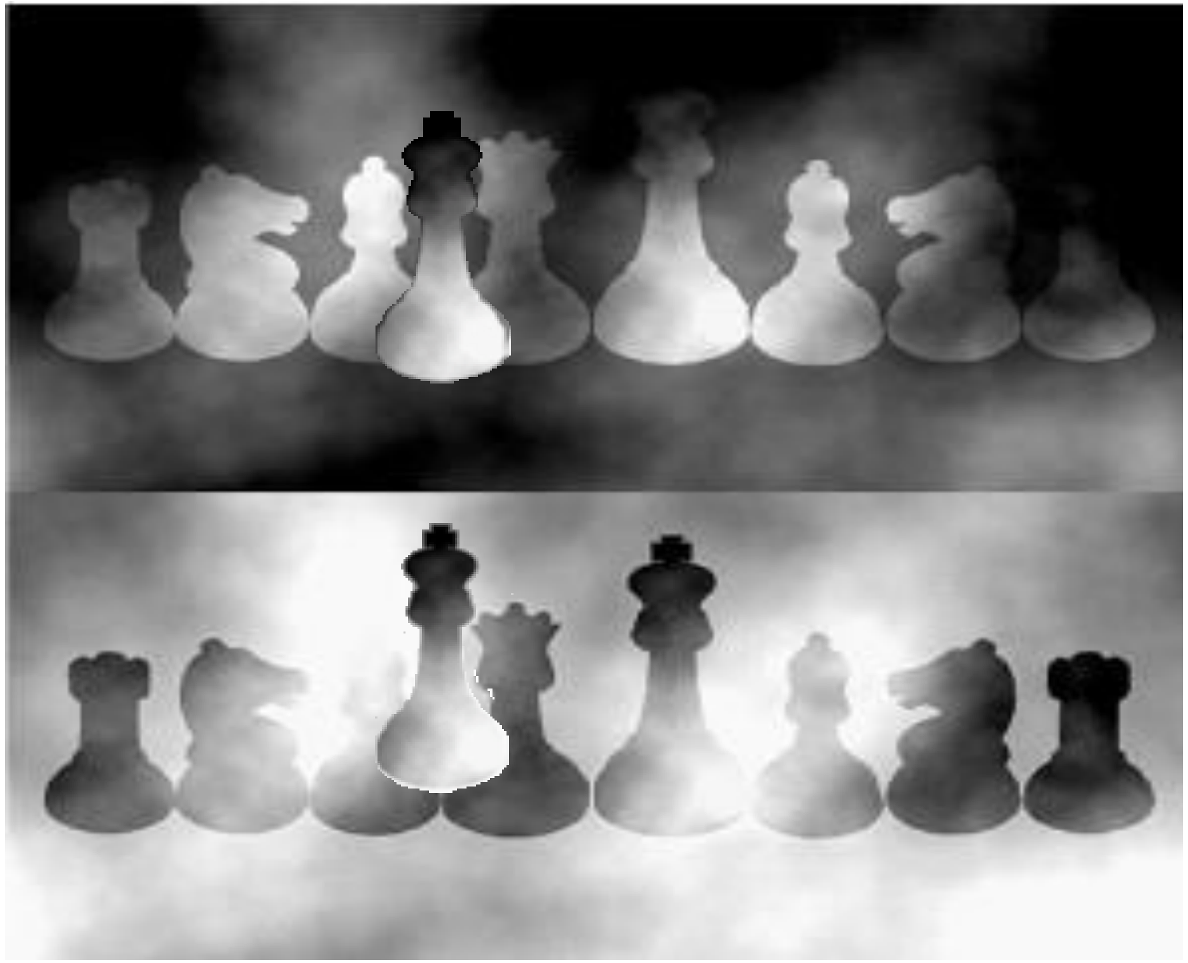
location
movement
spatial transformations
spatial relations

color
texture
pictorial detail
shape
size

Object processing

Ventral or "what" stream

Human Vision

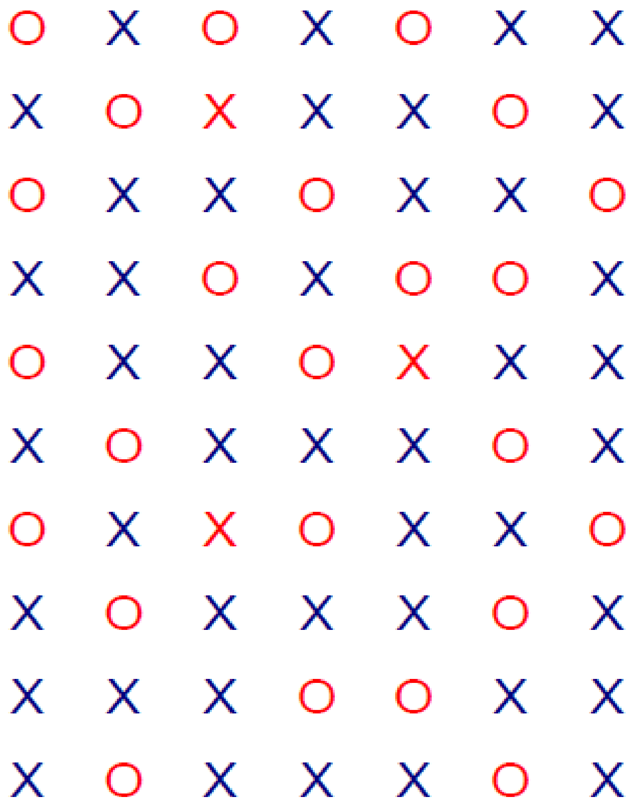


Human Vision

- Count the red crosses

Human Vision

- Count the red crosses



Human Vision

Human Vision

Human Vision

- Computer vision is much more difficult than you might initially think:
  - Your brain needs 25% of the cortex just to solve vision
  - The general computer vision problem is still largely unsolved
- Main problem: vision needs to deal with enormous variations in the signal
  - Some of these variations are relevant and others should be ignored

Gap between “pixels” and “meaning”

What we see

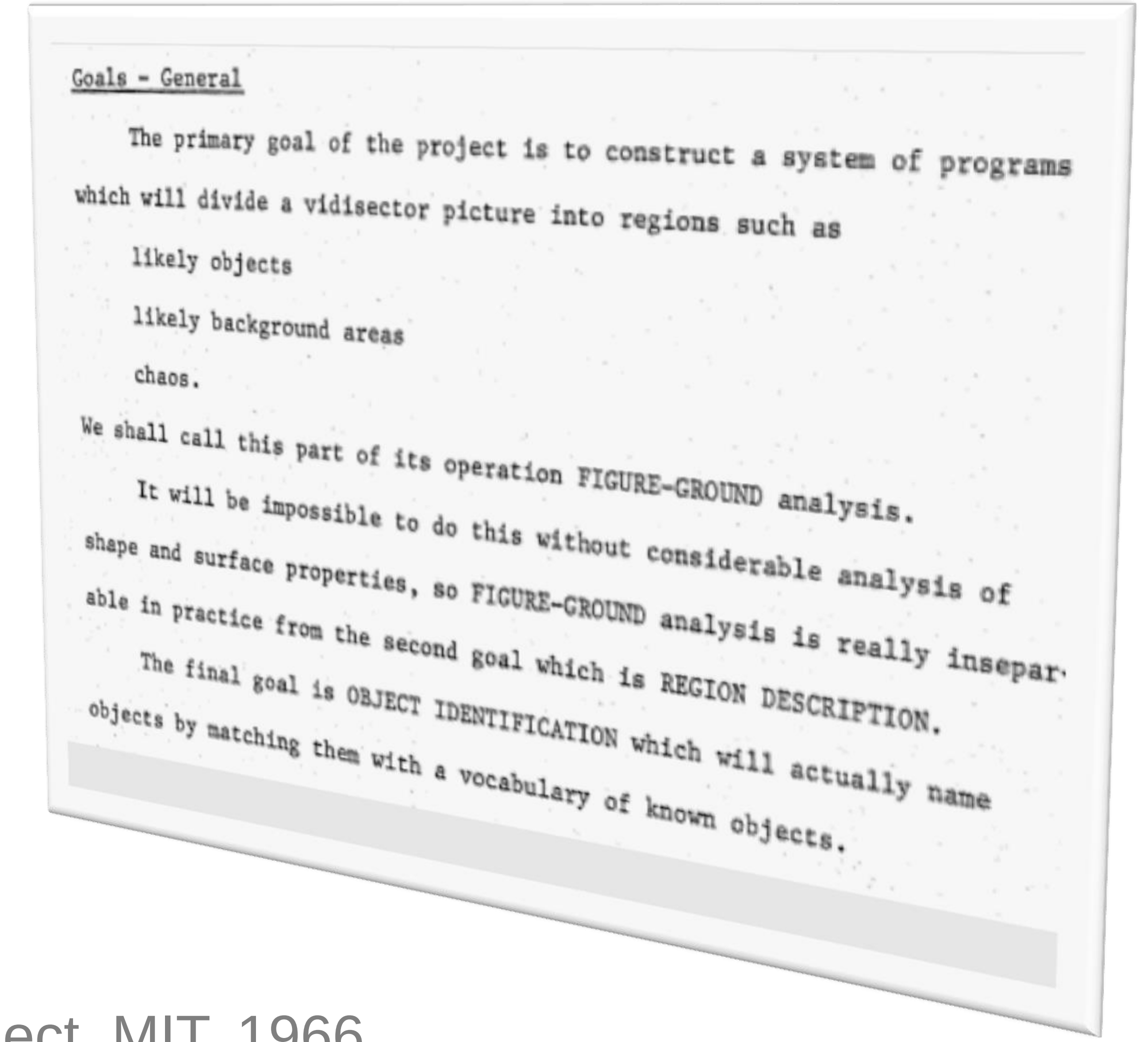
0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0

What a computer sees

Source: S. Narasimhan

Computer Vision: History

- An MIT undergraduate summer project*, in 1966, aimed to solve background/foreground segmentation and object detection/classification.
- It has been 57 years and we still work on the same problems.


Goals - General

The primary goal of the project is to construct a system of programs which will divide a vidisector picture into regions such as

likely objects

likely background areas

chaos.

We shall call this part of its operation FIGURE-GROUND analysis.

It will be impossible to do this without considerable analysis of shape and surface properties, so FIGURE-GROUND analysis is really inseparable in practice from the second goal which is REGION DESCRIPTION.

The final goal is OBJECT IDENTIFICATION which will actually name objects by matching them with a vocabulary of known objects.

*Seymour A Papert. The summer vision project, MIT, 1966.

Computer Vision

- Making useful decisions about real physical objects and scenes based on images (Shapiro & Stockman, 2001)
- Extracting descriptions of the world from pictures or sequences of pictures (Forsyth & Ponce, 2003)
- Analyzing images and producing descriptions that can be used to interact with the environment (Horn, 1986)
- Designing representations and algorithms for relating images to models of the world (Ballard & Brown, 1982)

Slide Credit: Rajesh Rao

Computer Vision

- How do we describe the variations within the *class* "chair"?

- Invariance to some variations can be obtained using hand-crafted models
- We generally try to *learn* invariance to the remaining variations from *examples*

Slide Credit: L. van der Maaten

Object Recognition

- Observation: chairs contain relatively lots of *edges*

Slide Credit: L. van der Maaten

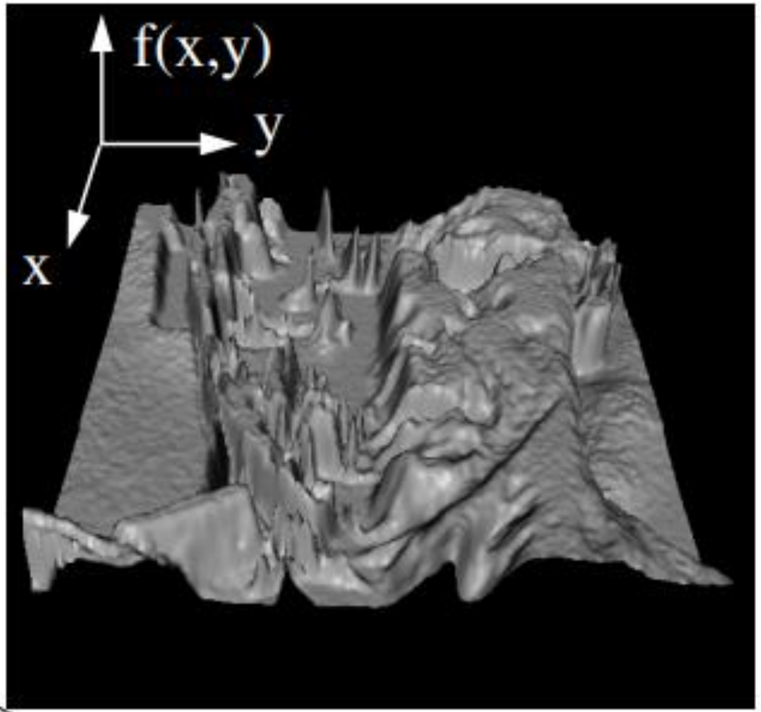
Basics: What is an image?

- Assume an image as a function, $f: \mathbb{R}^2 \rightarrow \mathbb{R}$:
  - $f(x, y)$ gives the intensity at position (x, y)
  - Realistically, an image is defined over a rectangle:
    - $f: [a, b] \times [c, d] \rightarrow [0,255]$

- Color image = Three functions combined together:
  - $f(x, y) = \begin{bmatrix} r(x, y) \\ g(x, y) \\ b(x, y) \end{bmatrix}$

Slide Credit: Rajesh Rao

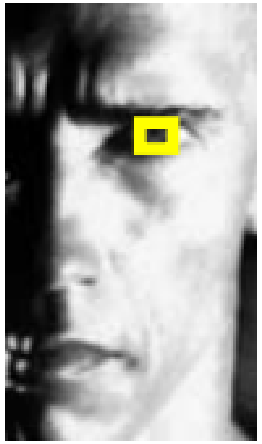
Basics: An image as a function



Bright regions are high, dark regions are low

Basics: Digital Images

- In computer vision we operate on digital (discrete) images:
  - Sample the 2D space on a regular grid
  - Quantize each sample (round to the nearest integer)
  - Each sample is a pixel (picture element)
  - If we assume each pixel as 1 byte, values range from 0 to 255



y

x

62	79	23	119	120	105	4	0
10	10	9	62	12	78	34	0
10	58	197	46	46	0	0	48
176	135	5	188	191	68	0	49
2	1	1	29	26	37	0	77
0	89	144	147	187	102	62	208
255	252	0	166	123	62	0	31
166	63	127	17	1	0	99	30

Slide Credit: Rajesh Rao

Basics: Preprocessing

- Range transformation (pixel processing):

$$g(x, y) = h\big(f(x, y)\big)$$

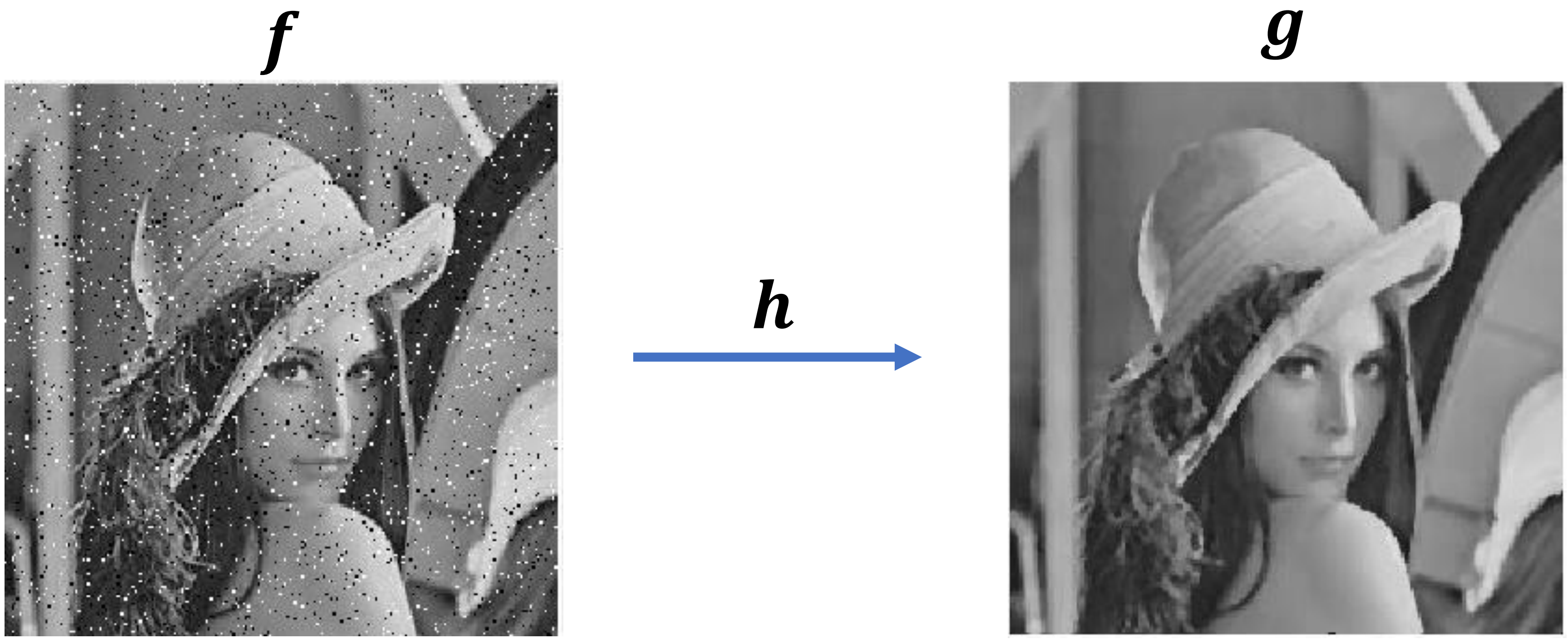
- Example: ?

Basics: Preprocessing

- Range transformation (pixel processing):

$$g(x, y) = h\big(f(x, y)\big)$$

- Example: Noise filtering



Basics: Preprocessing

- Domain transformation (geometric transform):

$$g(x,y) = f\left(h_x(x,y), h_y(x,y)\right)$$

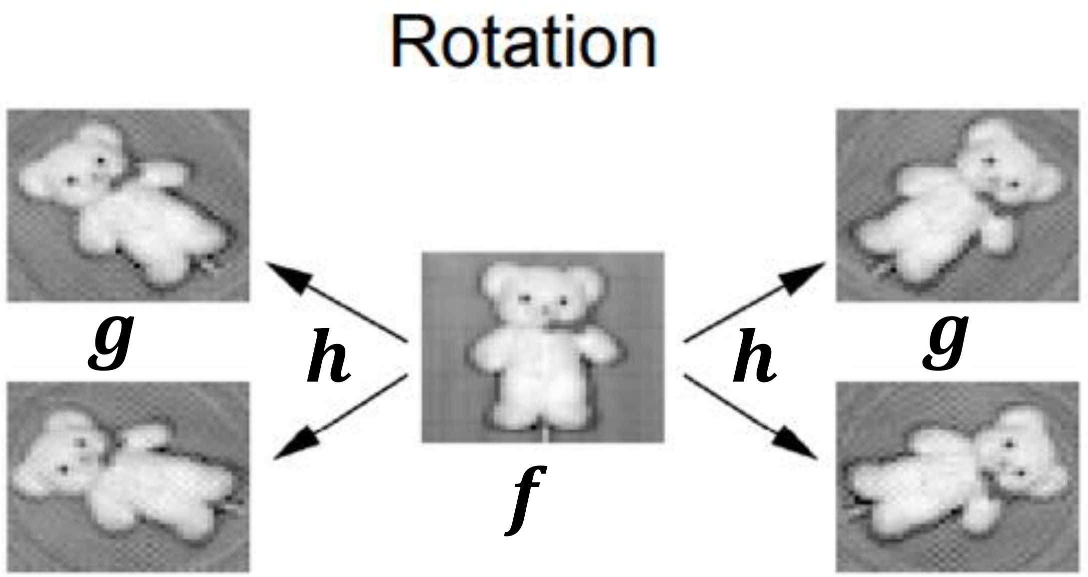
- Example: ?

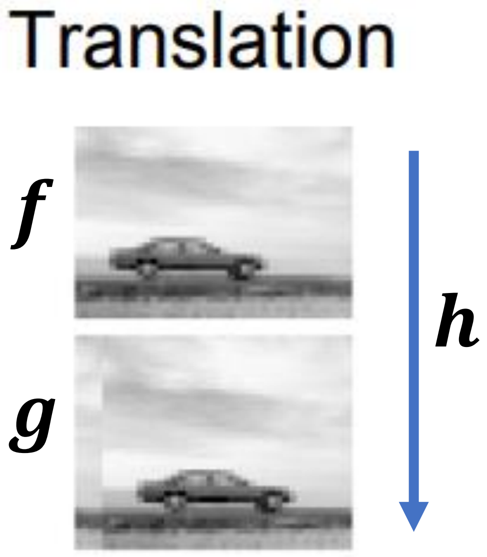
Basics: Preprocessing

- Domain transformation (geometric transform):

$$g(x,y) = f\big(h_x(x,y), h_y(x,y)\big)$$

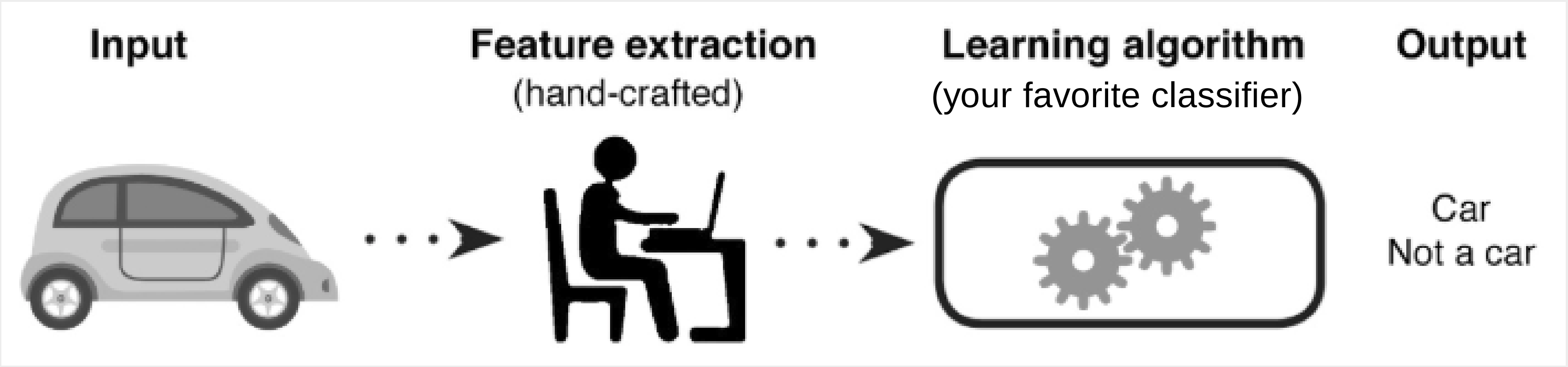
- Example: Rotation / Translation

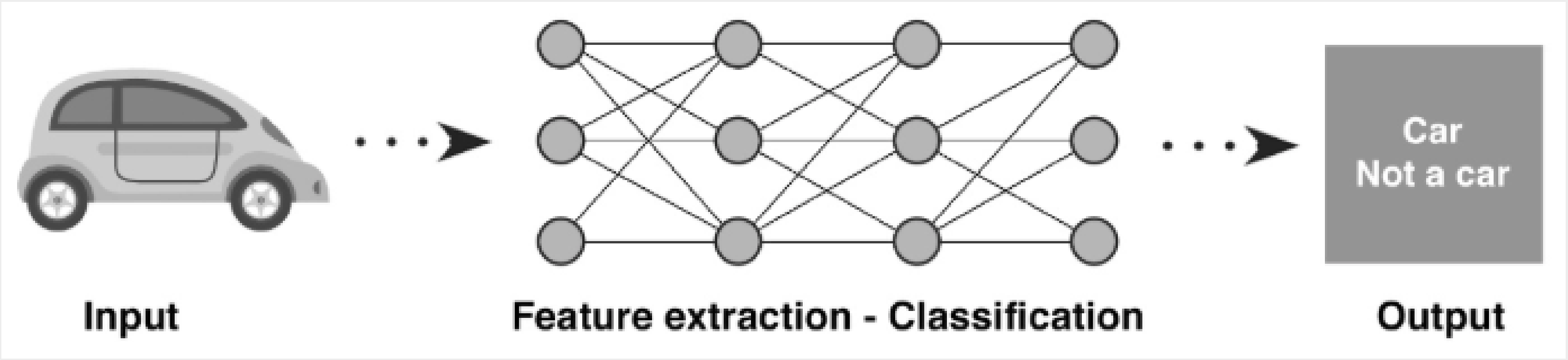




Slide Credit: Rajesh Rao

Basics: Analysis Pipeline



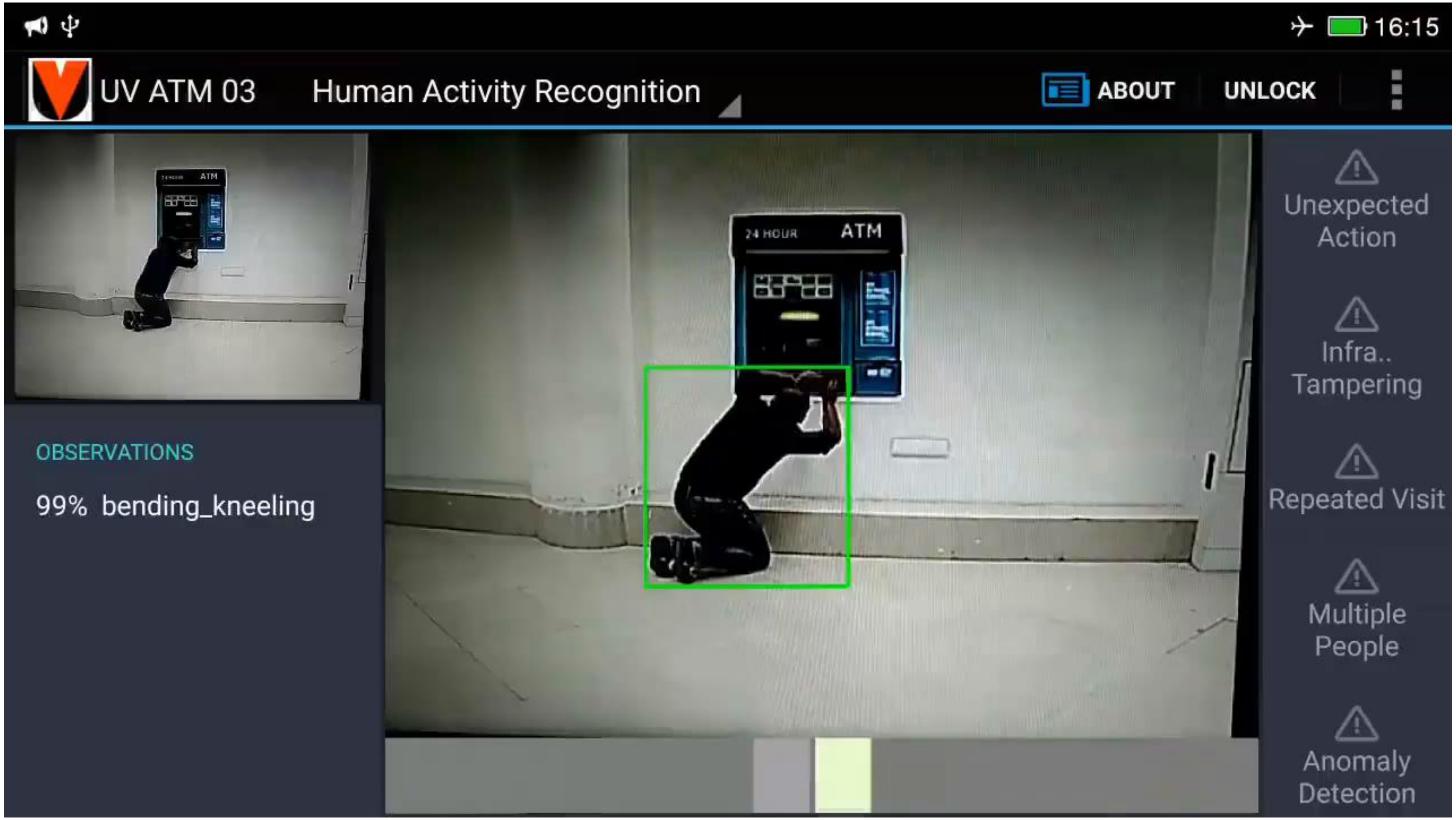


Surveillance

- Face Recognition
- Object Detection
- Tracking

Surveillance

- Anomaly Detection
- Action & Activity Recognition



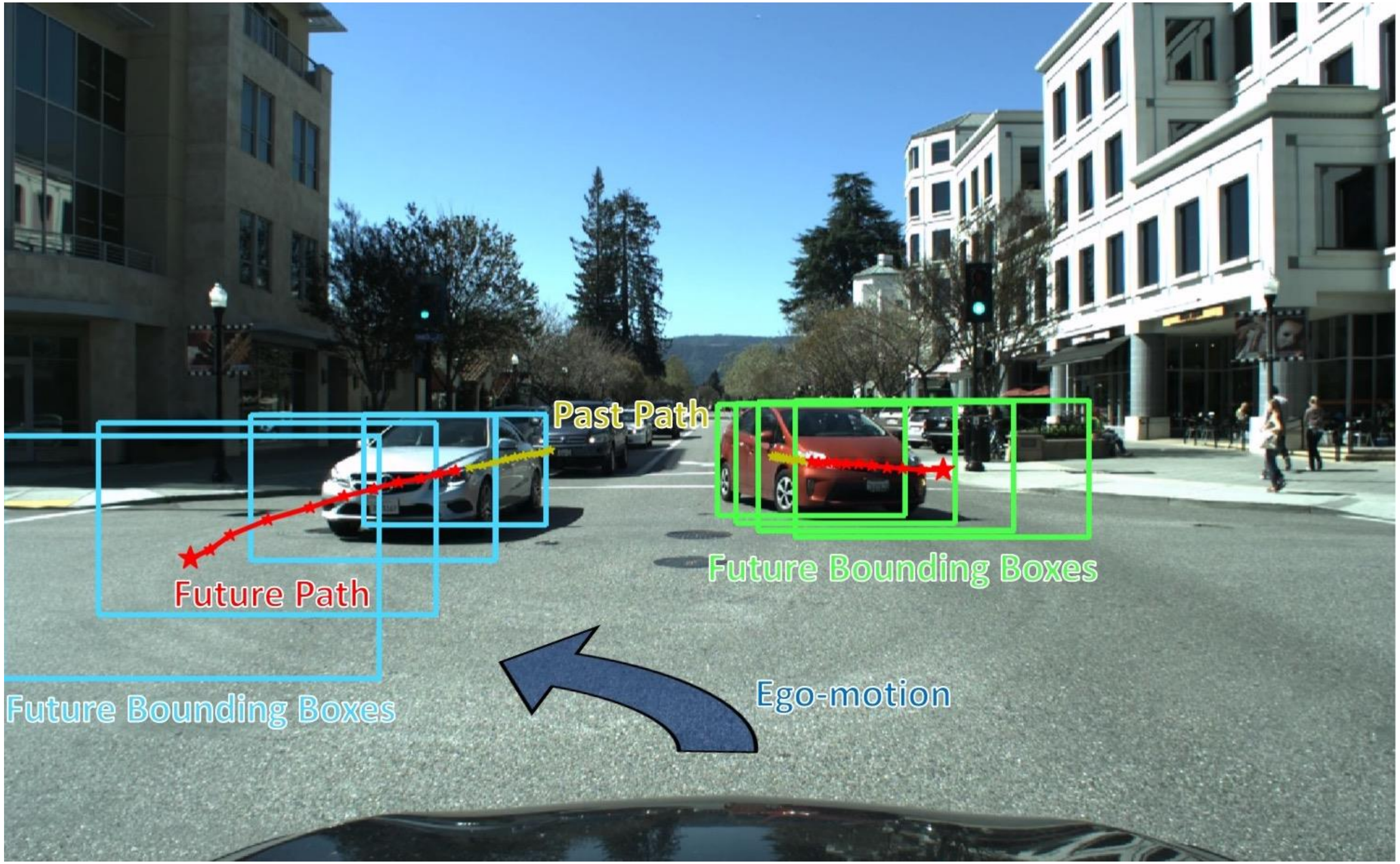
Autonomous Vehicles

Autonomous Vehicles

- Detection & Segmentation
- Tracking

Autonomous Vehicles

- Future Prediction

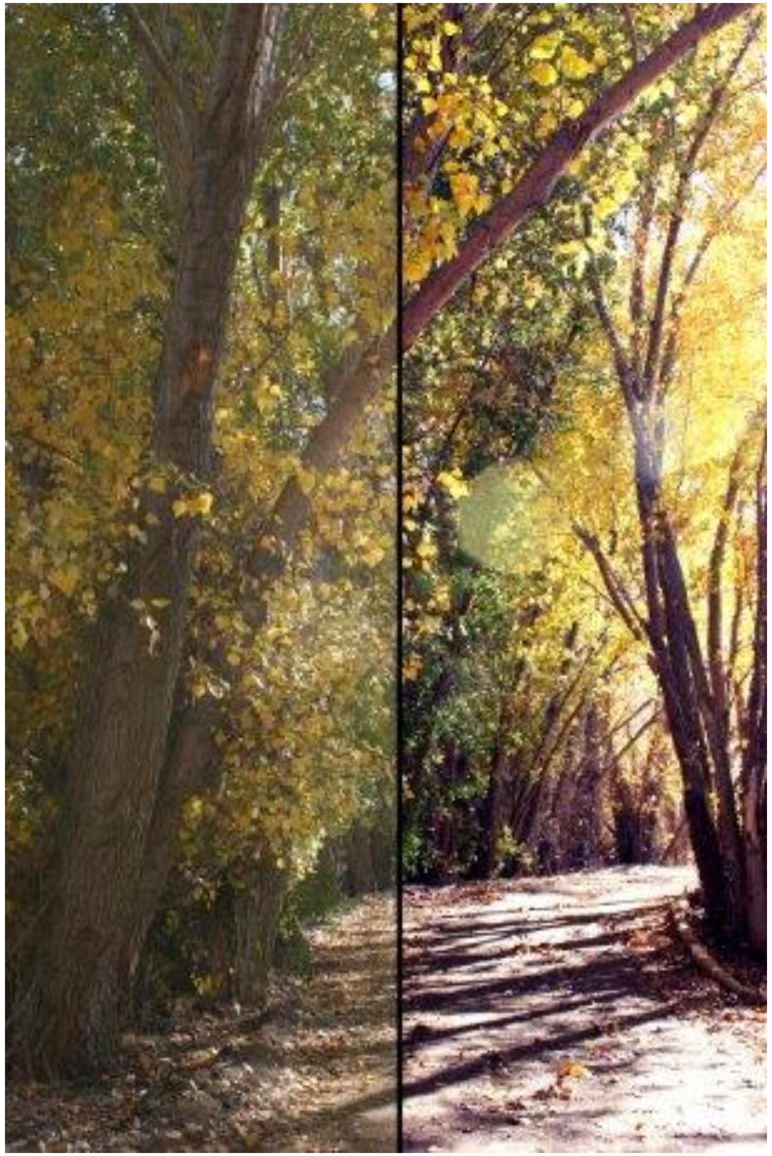


Autonomous Vehicles

- Gaze Estimation
- Mood Recognition
- Fatigue Monitoring

Photography

- Face Detection / Smile Shutter
- Focus Tracking
- Image Stabilization
- Color Enhancement

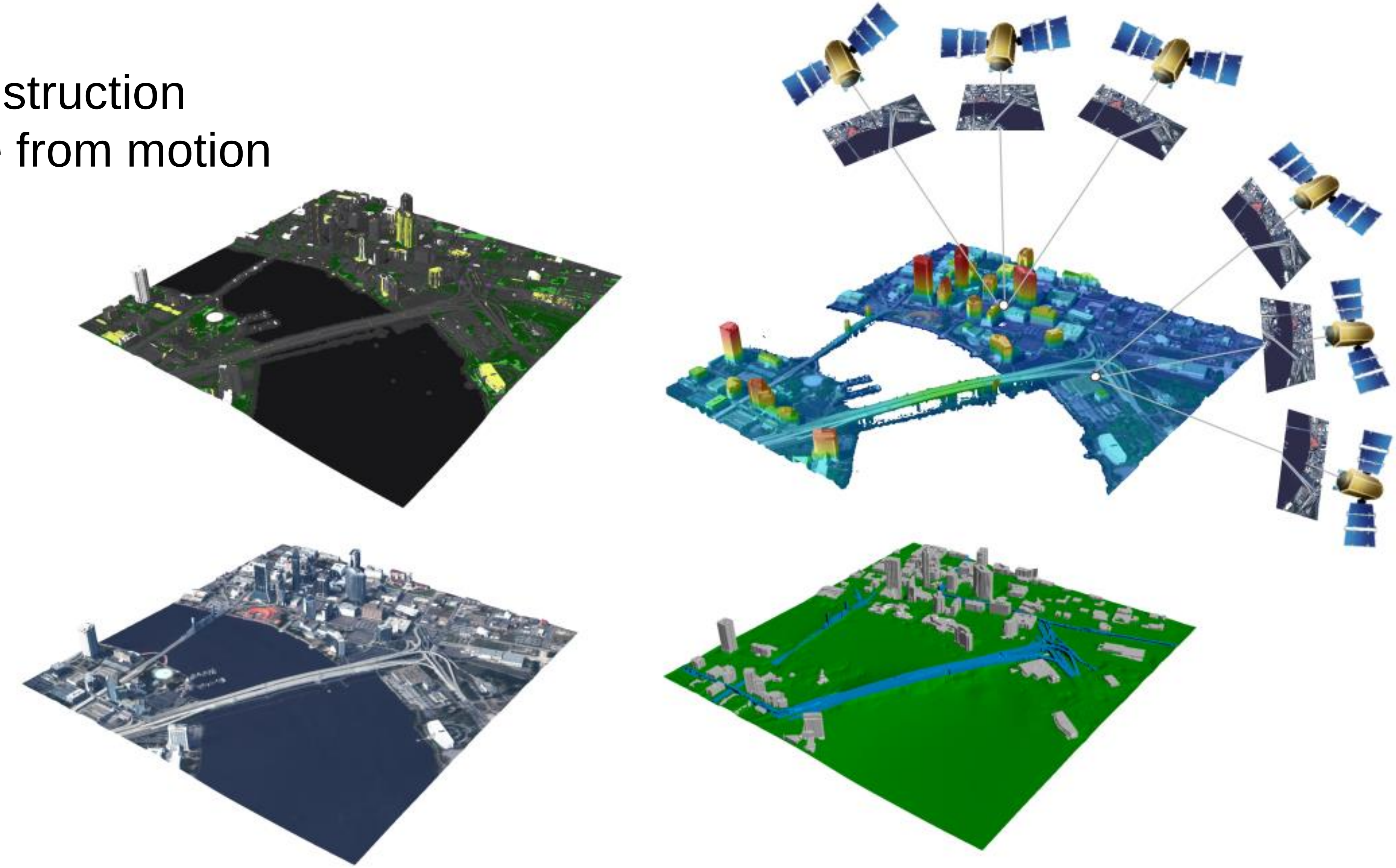


Special Effects

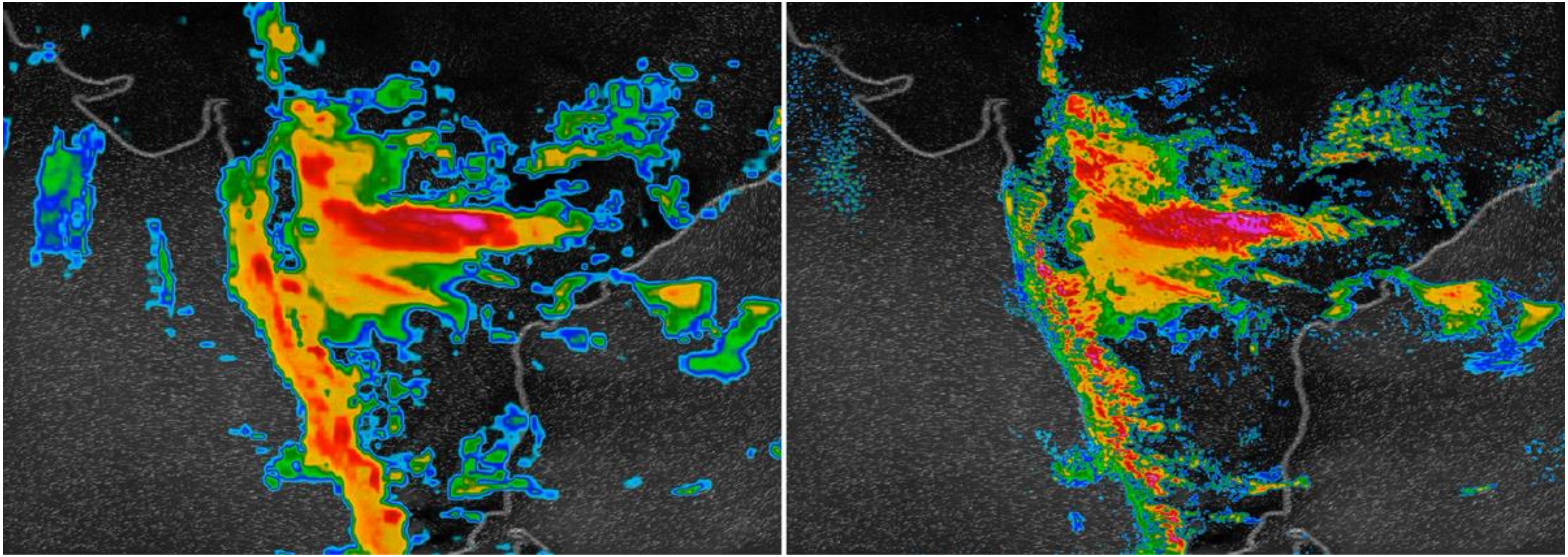
Special Effects (3D)

Map & 3D Model Generation

- 3D reconstruction
- Structure from motion



Weather Forecasting



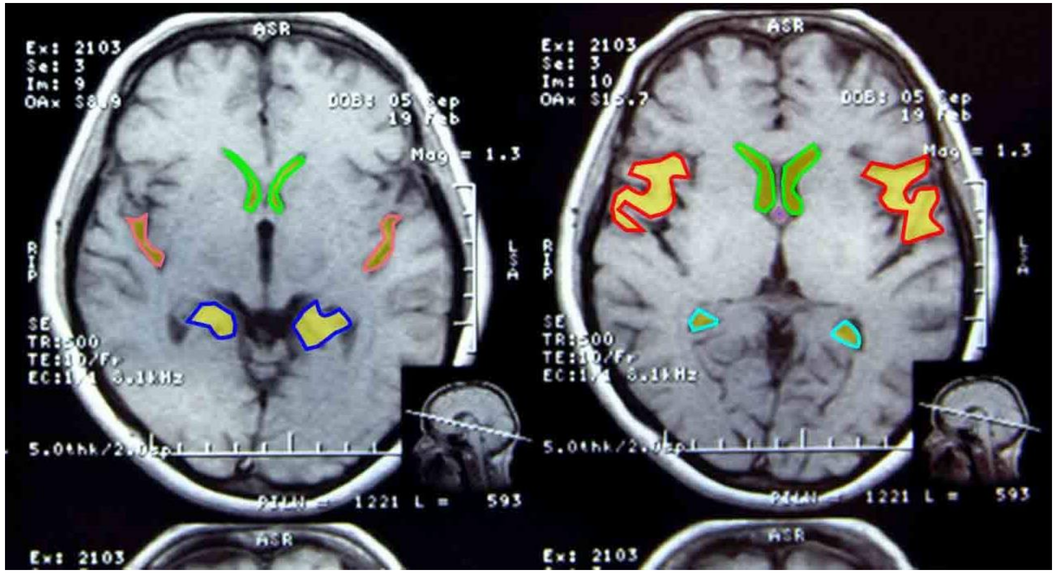
- Spectral image analysis
- Image segmentation

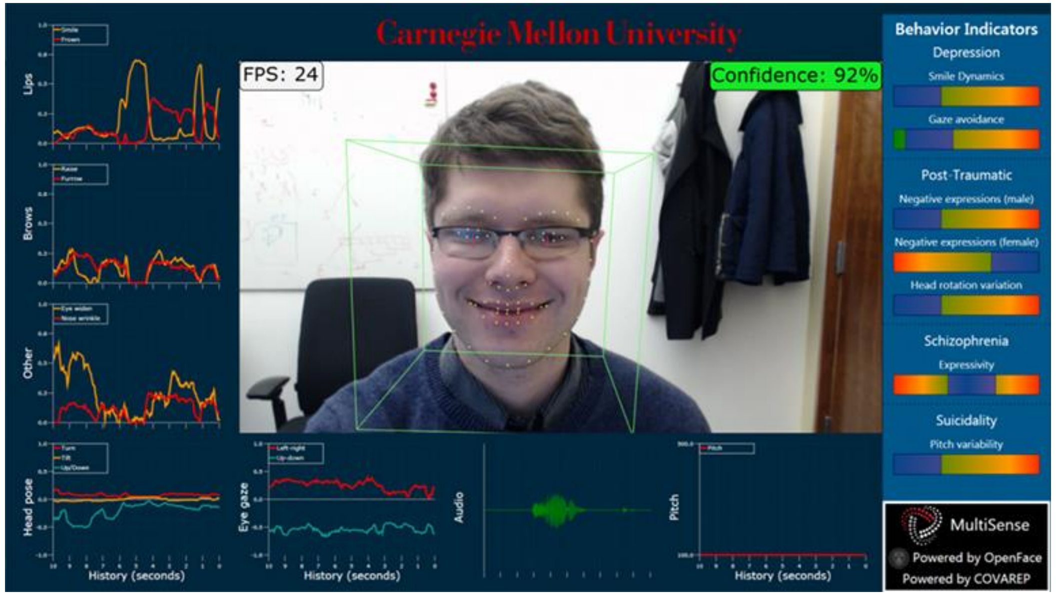
Robotics

Quality Control & Monitoring

Healthcare

- Facial expression analysis
- Pose estimation
- Medical image analysis
- Motion magnification
- Subtle motion tracking



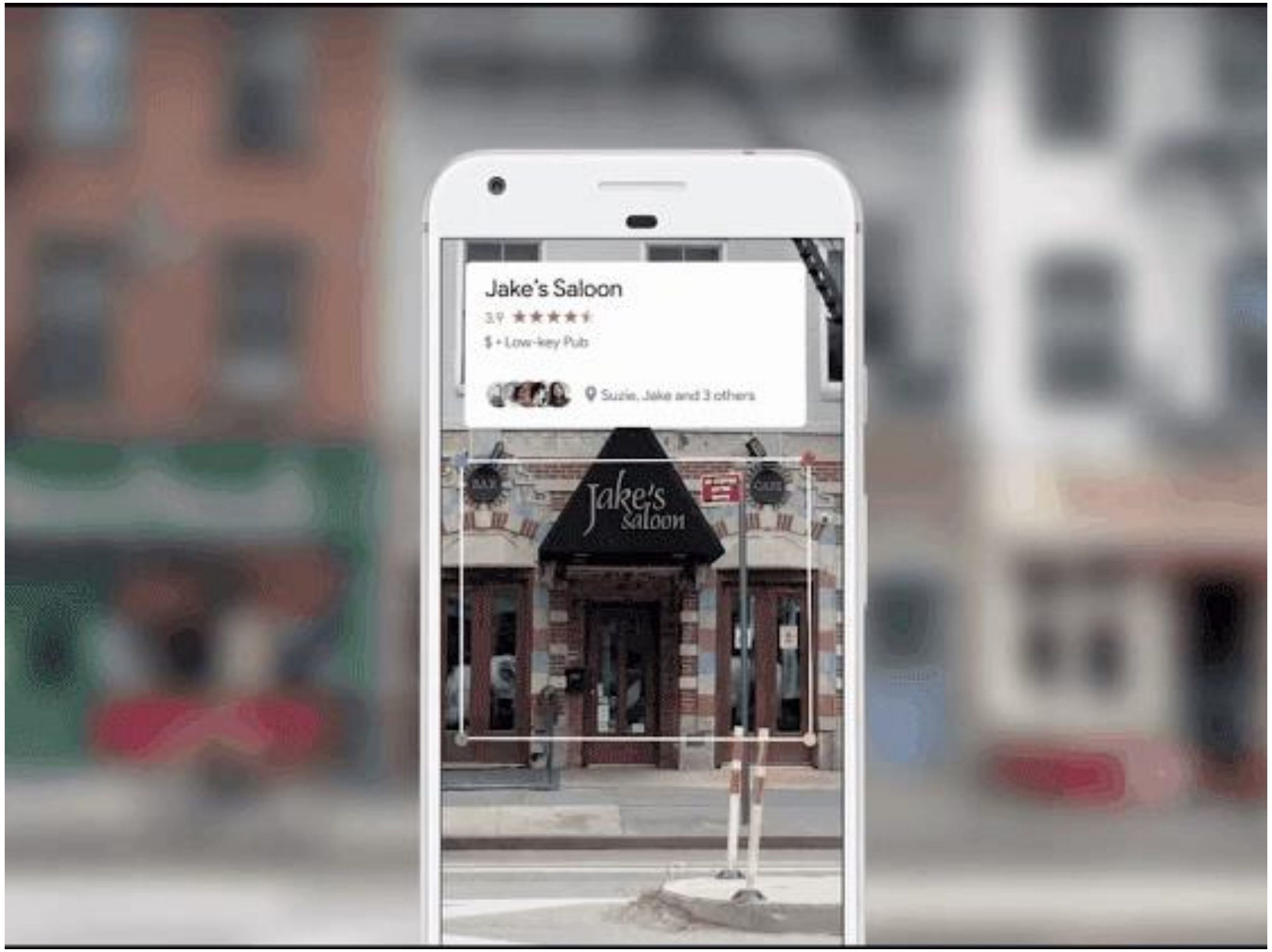


Neuromarketing

- Gaze estimation / tracking
- Facial emotion recognition
- Age & gender estimation
- Appreciation recognition

Visual Search

- https://lens.google.com/

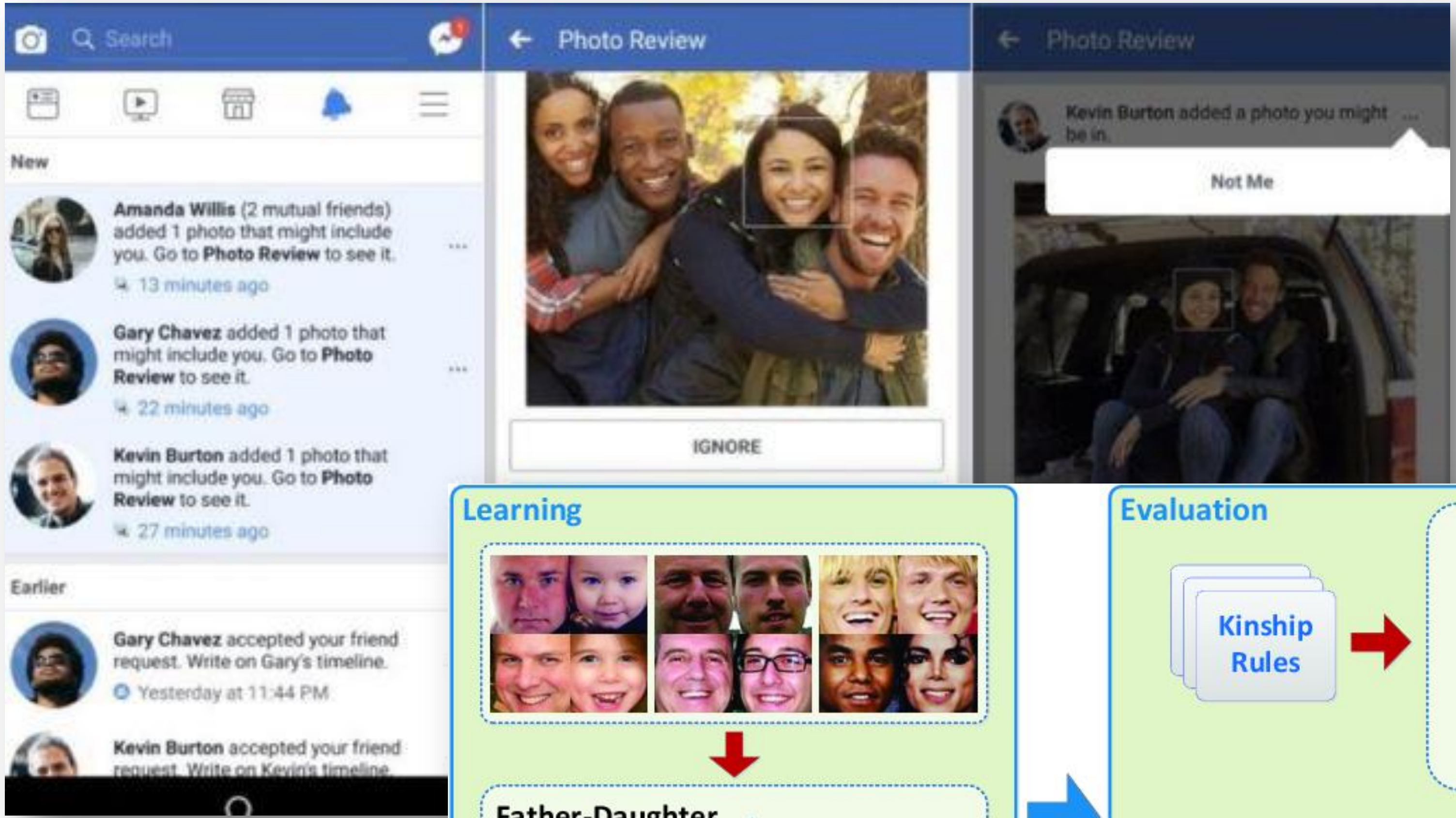


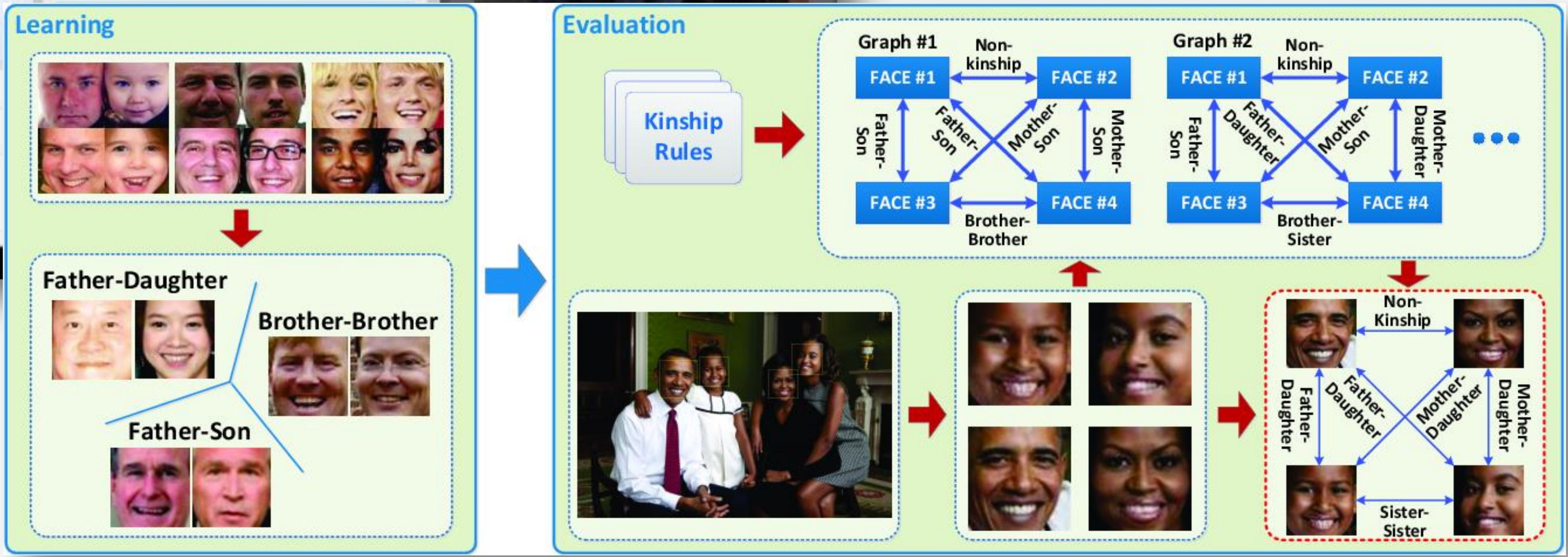
Gaming / Human Computer Interaction

- Pose and motion tracking
- Gesture recognition

Sports Analytics

Image Analytics: Identity Recognition / Tagging





- Face detection
- Face recognition
- Kinship recognition

Image Analytics: Captioning

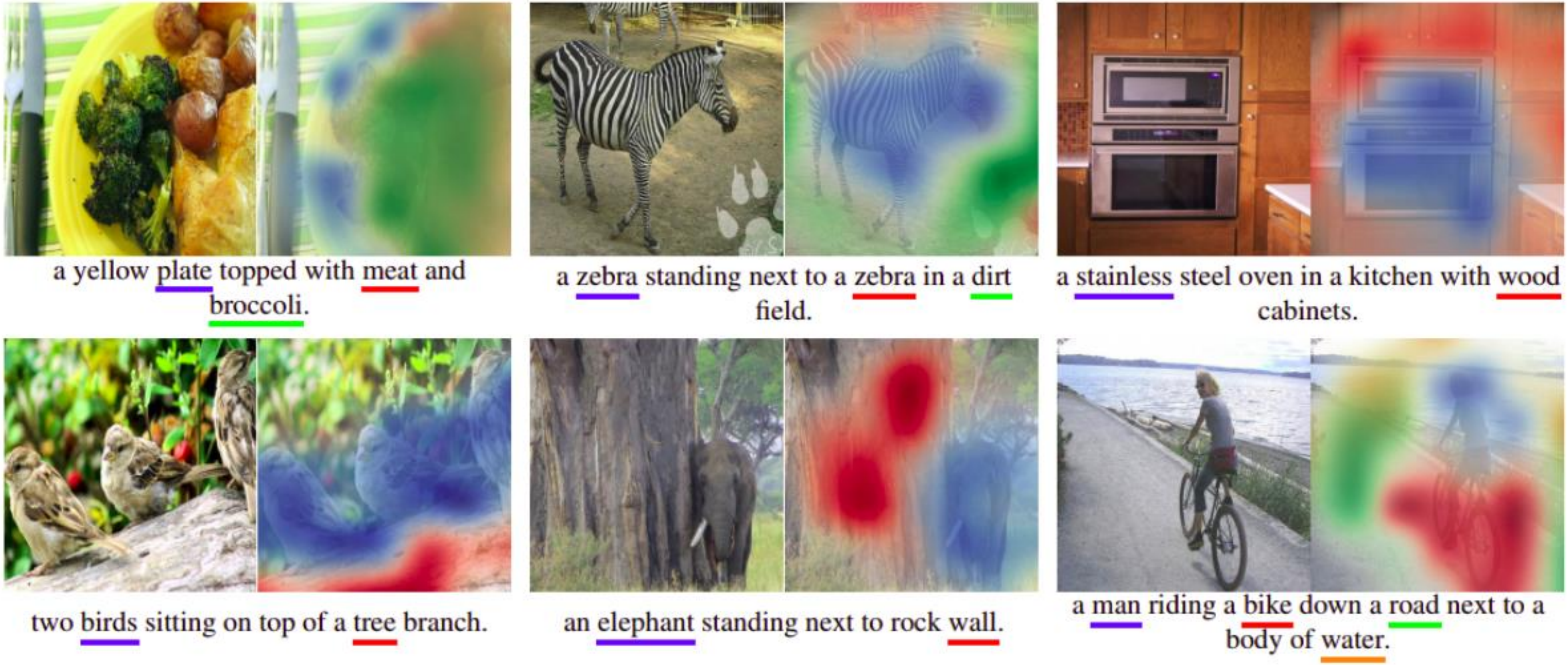
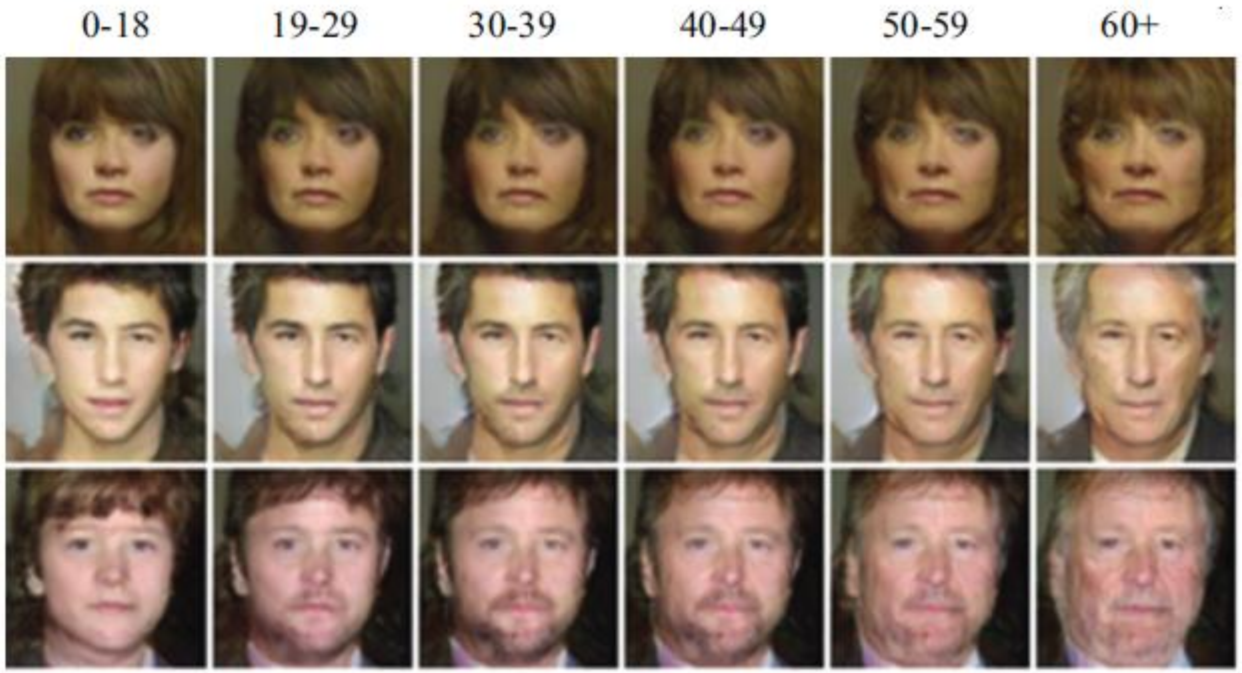
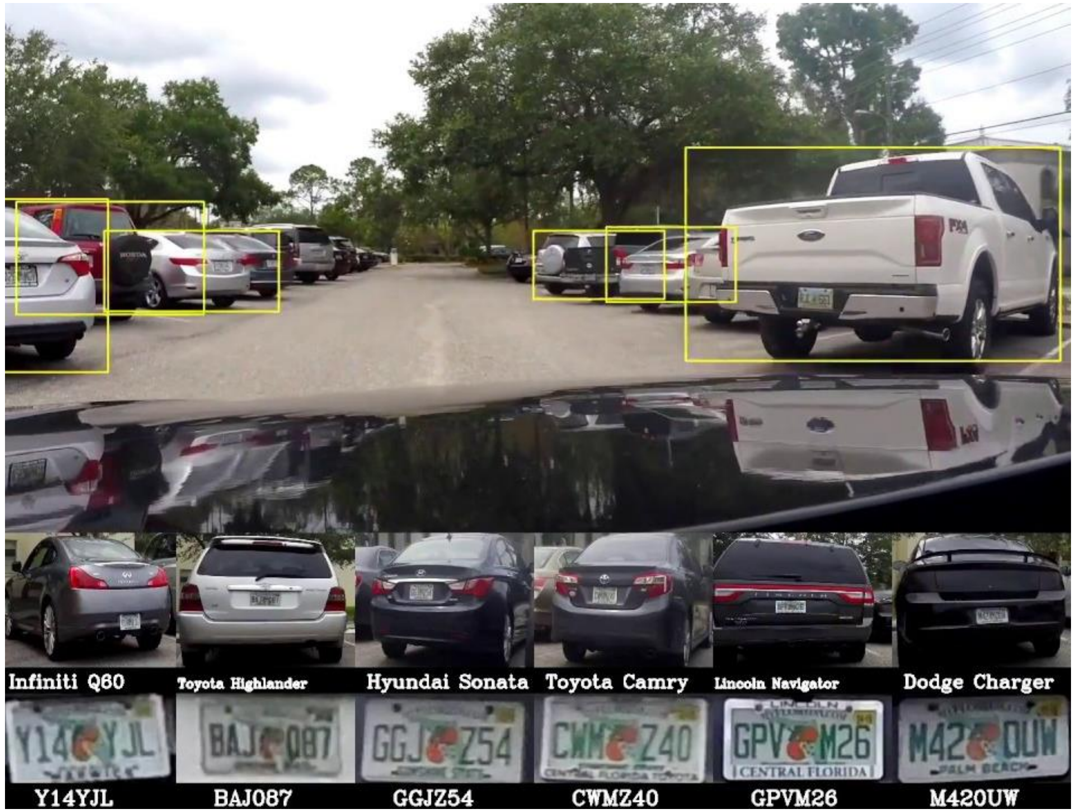


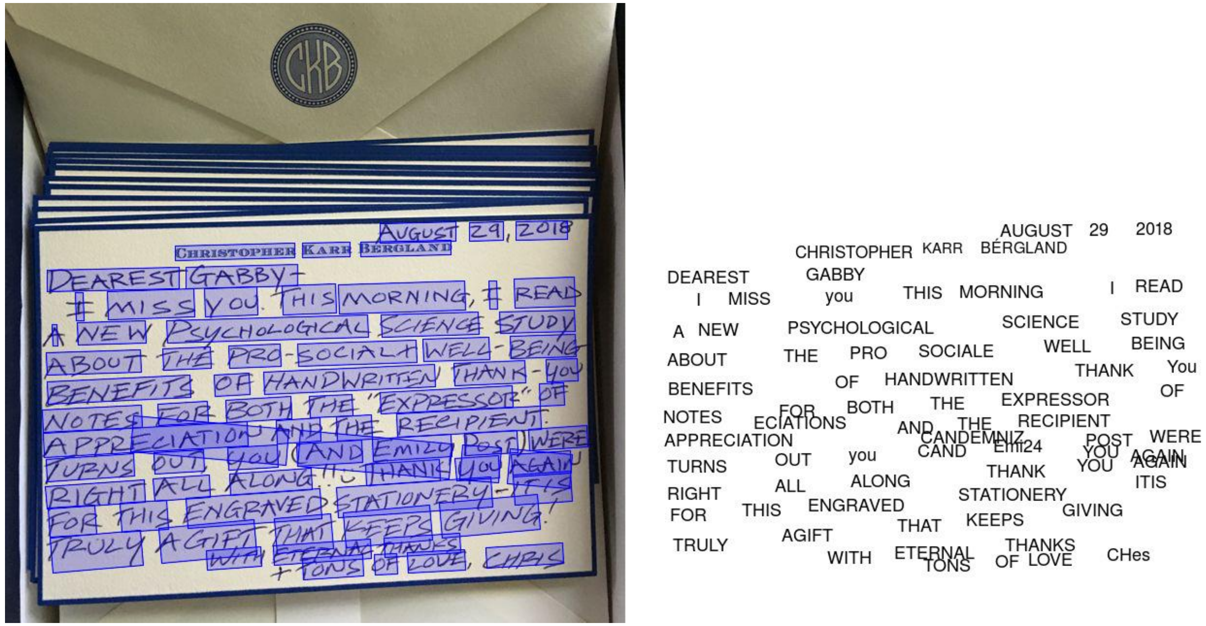
Image Synthesis/Modification



Style Transfer

Text Recognition / Document Digitization





And many more applications…