



GE 461

Introduction to Data Science

Spring 2024

Applications: Computer Vision

Hamdi Dibekliolu

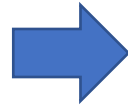
Computer Vision

*According
to
Hollywood*



Computer Vision

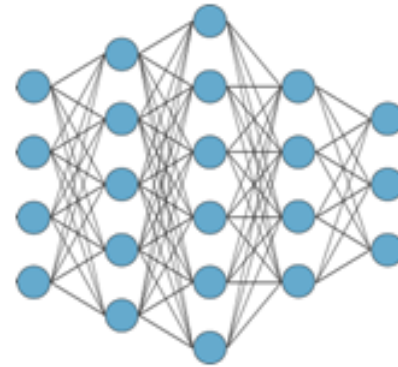
Scene



Sensor



Interpreter

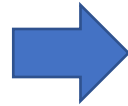


Interpretation

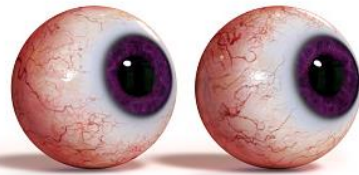
Mountain
Lake
Trees
Sky

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

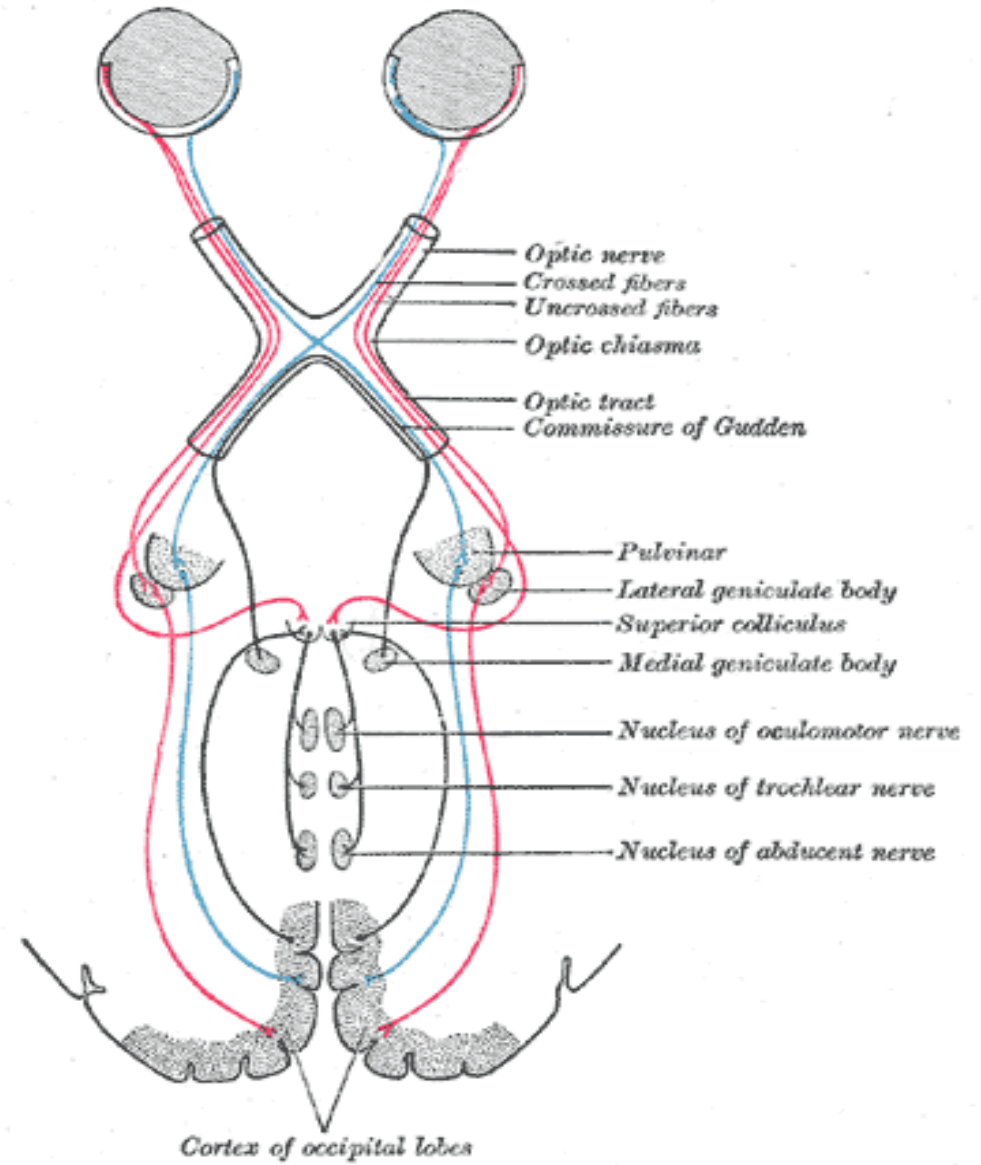
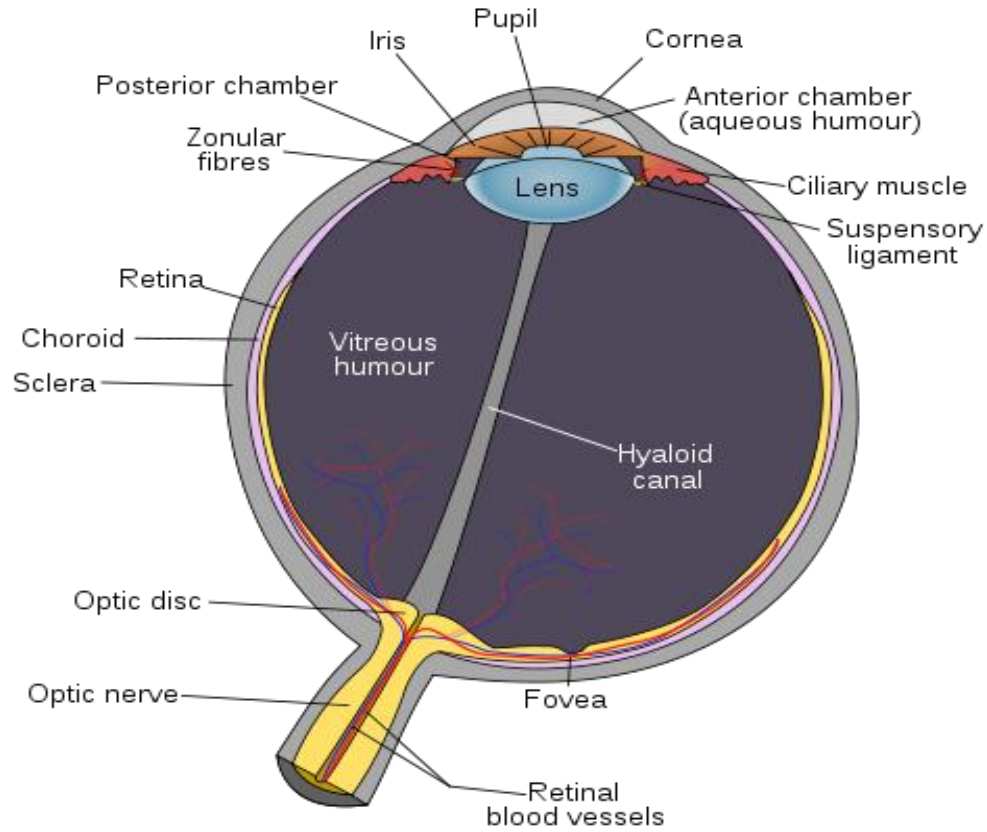
Human Vision

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

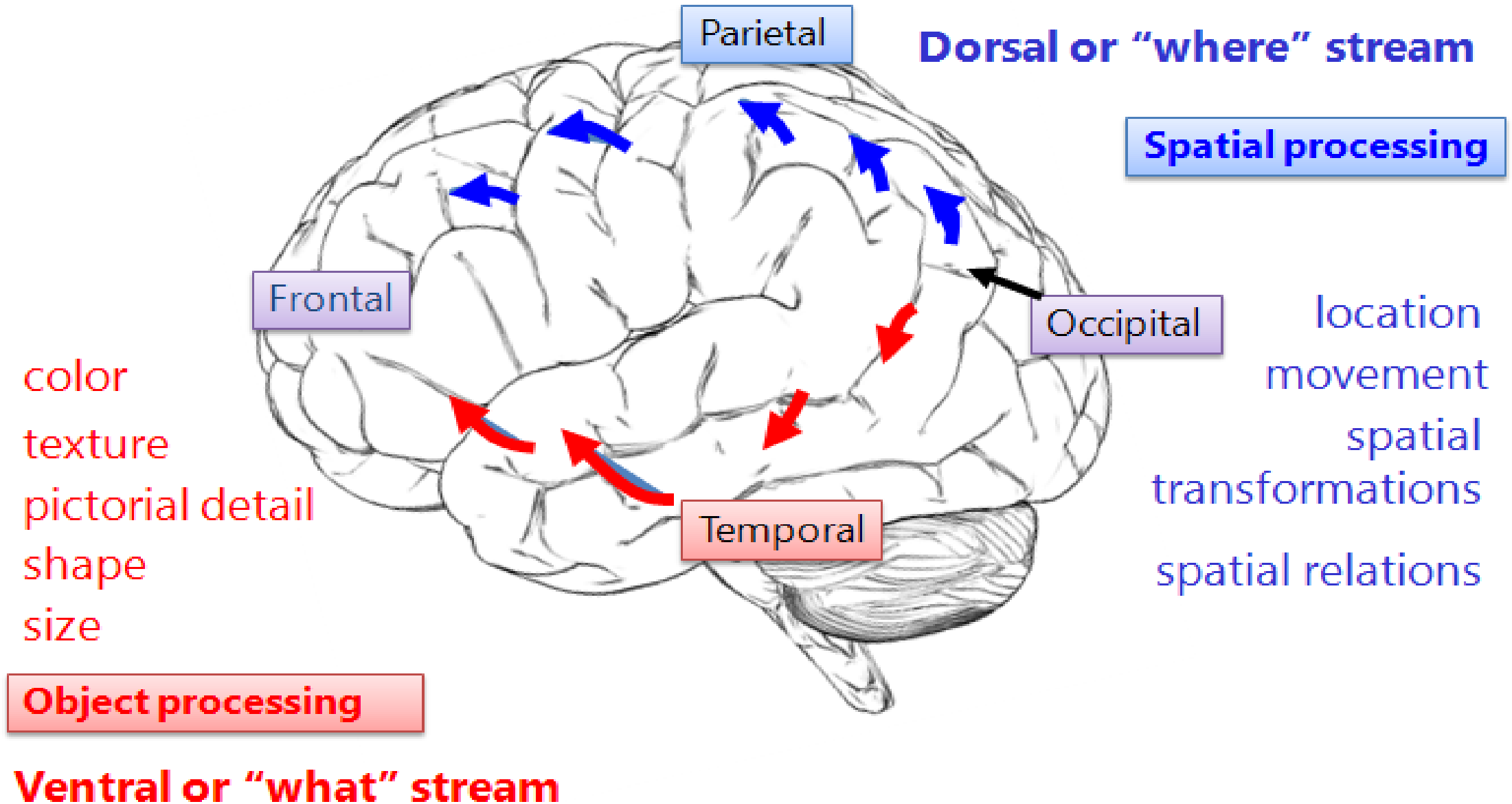
Human Vision

- 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

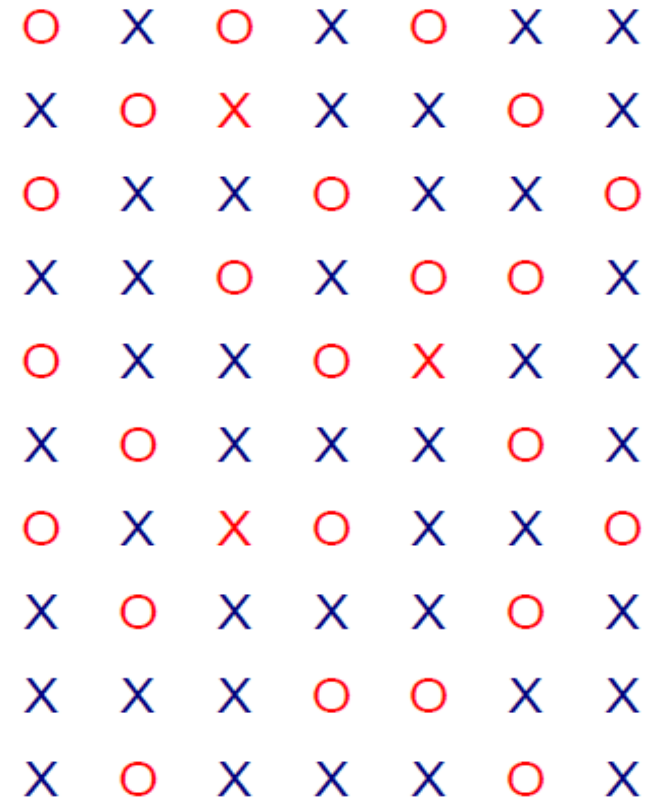


Human Vision



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”



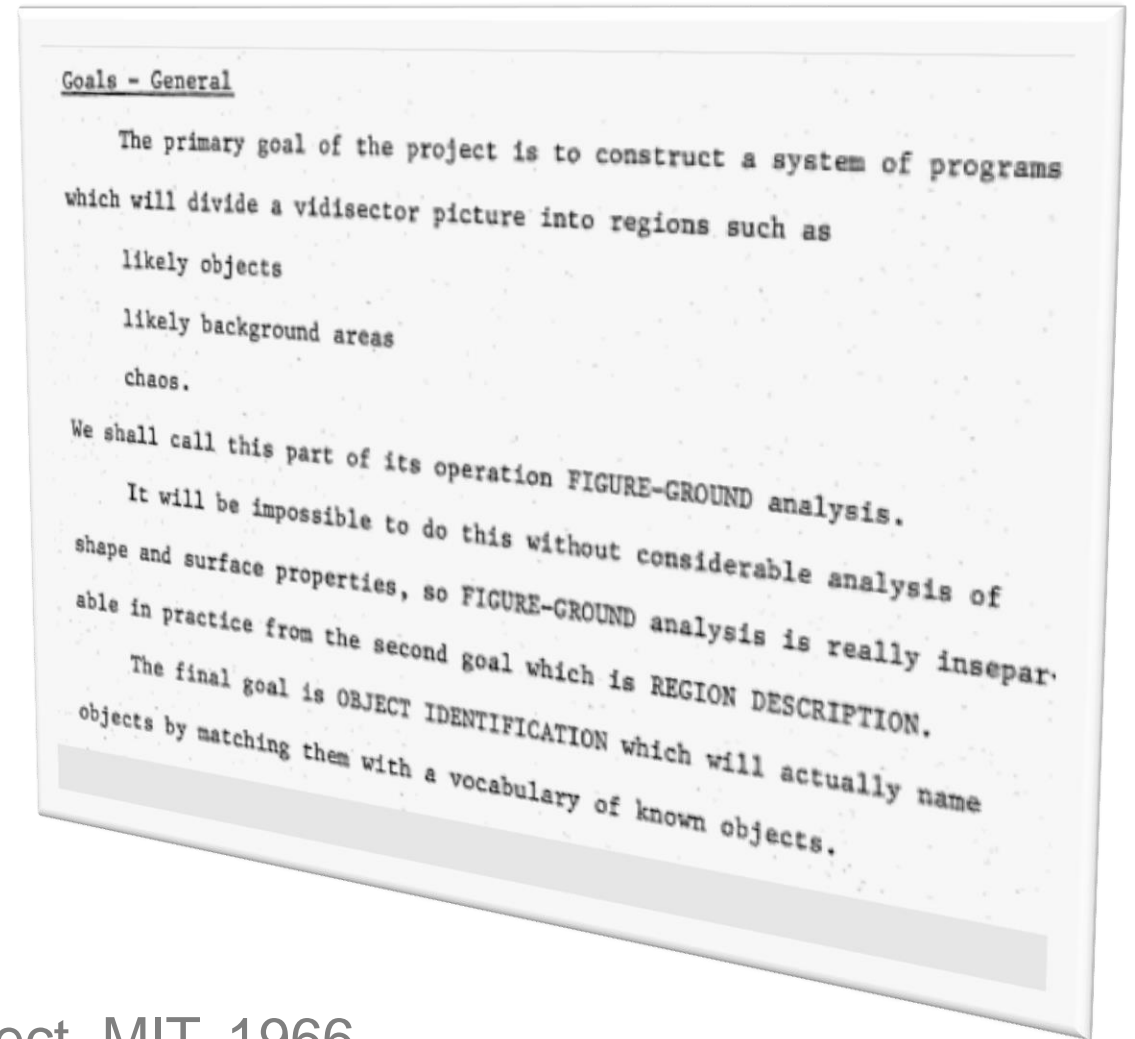
What we see

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

Computer Vision: History

- An MIT undergraduate summer project*, in 1966, aimed to solve background/foreground segmentation and object detection/classification.
- It has been 58 years and we still work on the same problems.



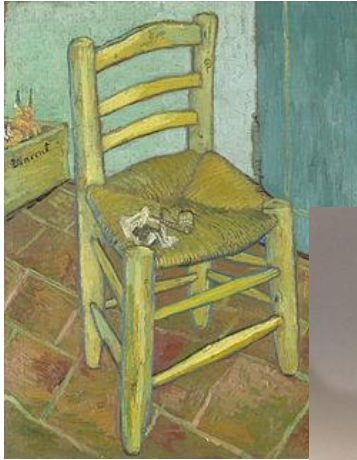
*Seymour A Papert. The summer vision project, MIT, 1966.

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)

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*

Object Recognition

- Observation: chairs contain relatively lots of *edges*



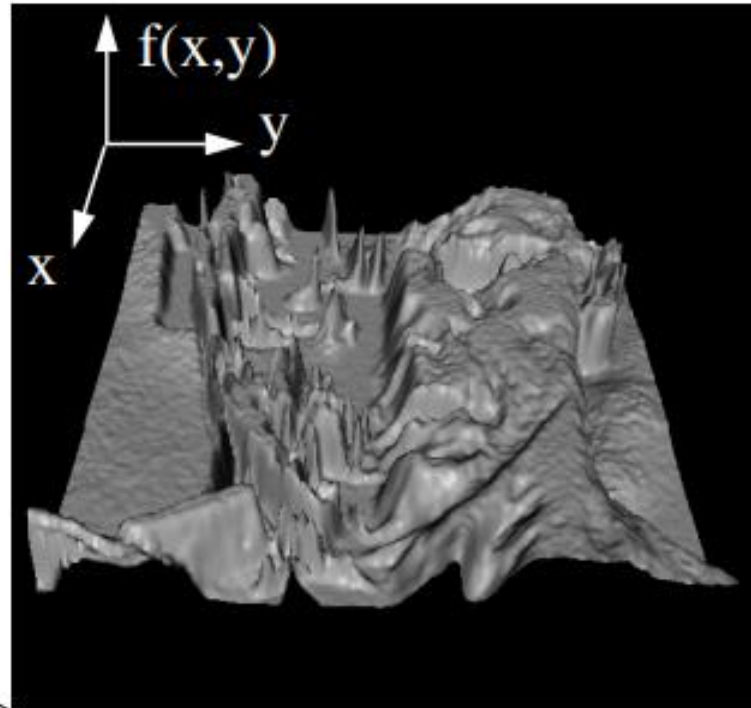
Basics: What is an image?

- 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:

- $f(x, y) = \begin{bmatrix} r(x, y) \\ g(x, y) \\ b(x, y) \end{bmatrix}$

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



Diagram illustrating a 2D grid of pixels with axes labeled x (vertical) and y (horizontal).

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

Basics: Preprocessing

- Range transformation (pixel processing):

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

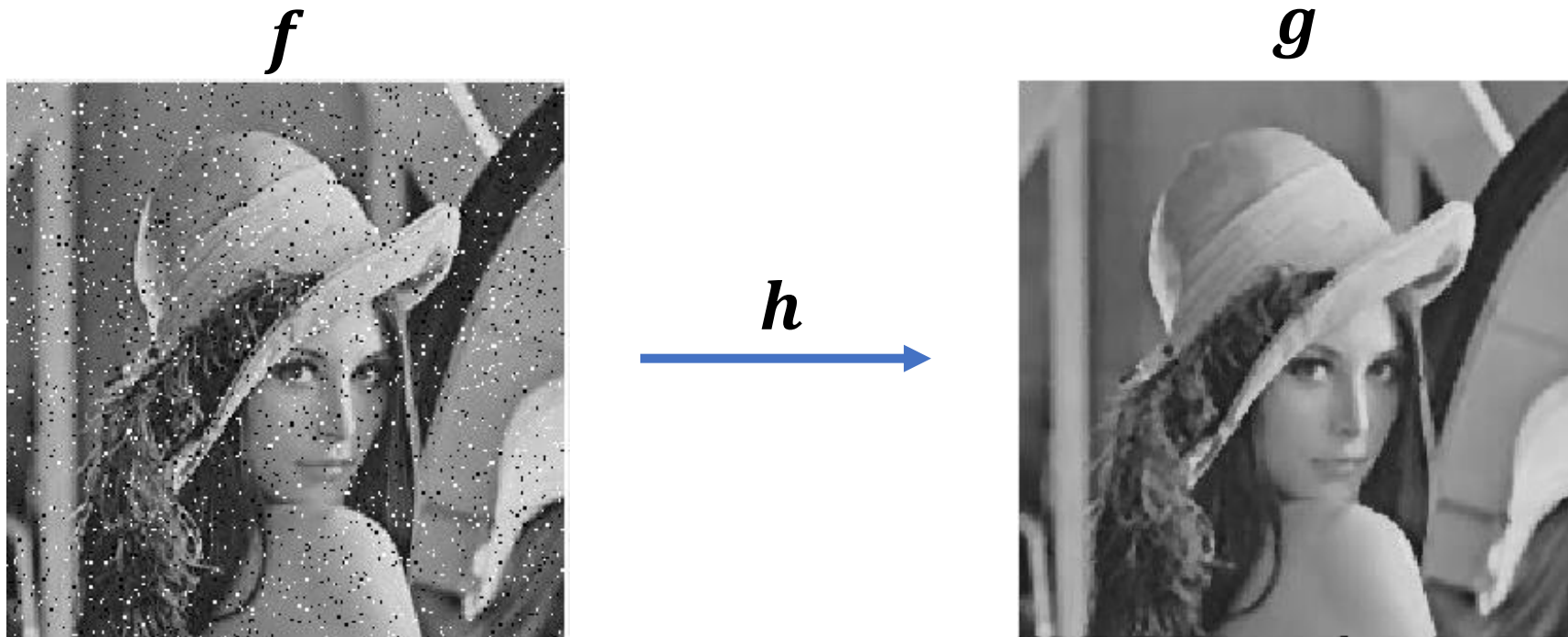
- Example: ?

Basics: Preprocessing

- Range transformation (pixel processing):

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

- Example: Noise filtering



Basics: Preprocessing

- Domain transformation (geometric transform):

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

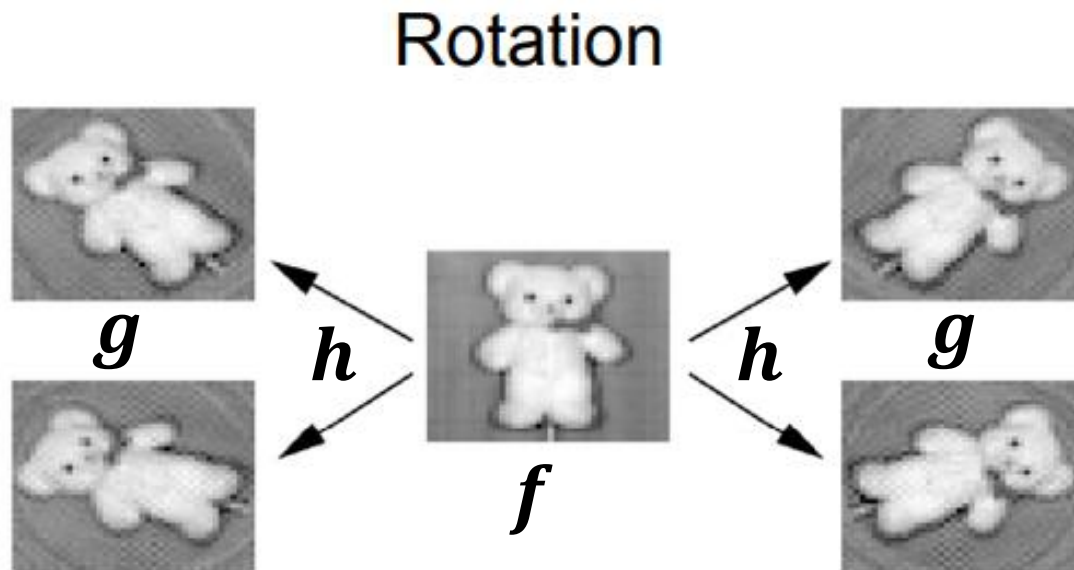
- Example: ?

Basics: Preprocessing

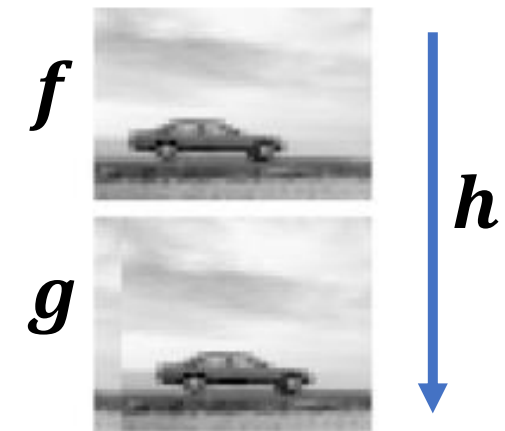
- Domain transformation (geometric transform):

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

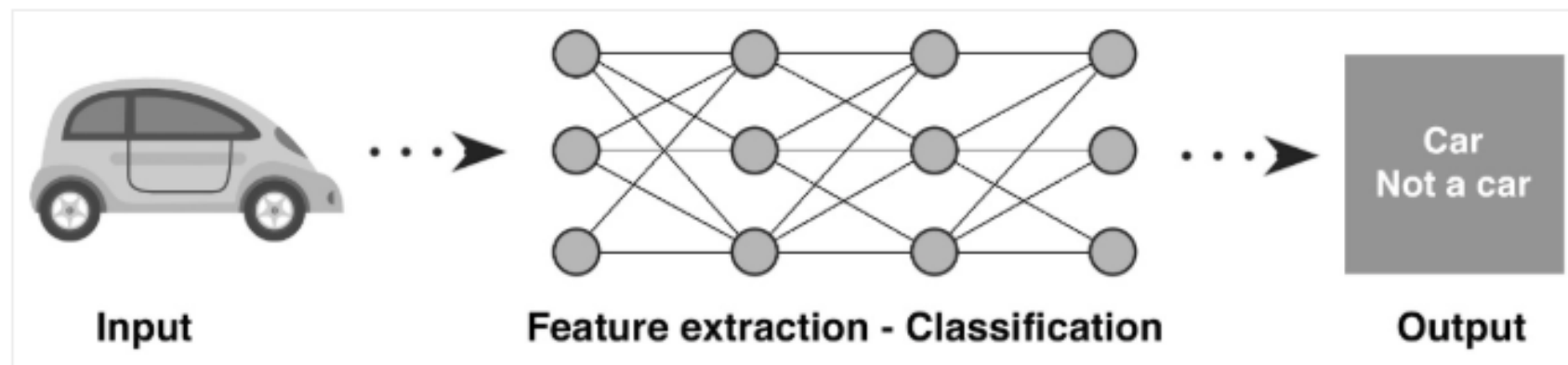
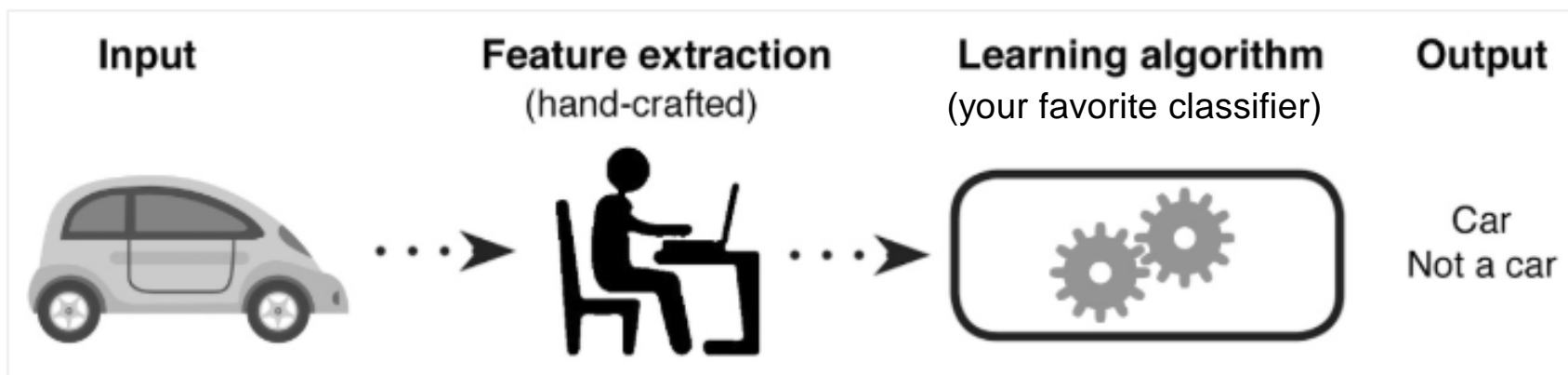
- Example: Rotation / Translation



Translation



Basics: Analysis Pipeline



Surveillance

- Face Recognition
- Object Detection
- Tracking



Surveillance

- Anomaly Detection
- Action & Activity Recognition

The screenshot displays a mobile application interface for human activity recognition. At the top, the status bar shows a signal strength icon, a USB icon, and the time 16:15. Below the status bar, the app header includes a logo, the text "UV ATM 03", and "Human Activity Recognition". To the right of the header are buttons for "ABOUT" and "UNLOCK", and a menu icon. The main content area is split into three sections: a small thumbnail on the left showing a person kneeling by an ATM, a large central video feed of the same scene with a green bounding box around the person, and a vertical list of detection alerts on the right. The alerts include "Unexpected Action", "Infra.. Tampering", "Repeated Visit", "Multiple People", and "Anomaly Detection". At the bottom left, an "OBSERVATIONS" section shows "99% bending_kneeling".

UV ATM 03 Human Activity Recognition

ABOUT UNLOCK

Unexpected Action

Infra.. Tampering

Repeated Visit

Multiple People

Anomaly Detection

OBSERVATIONS

99% bending_kneeling

Autonomous Vehicles



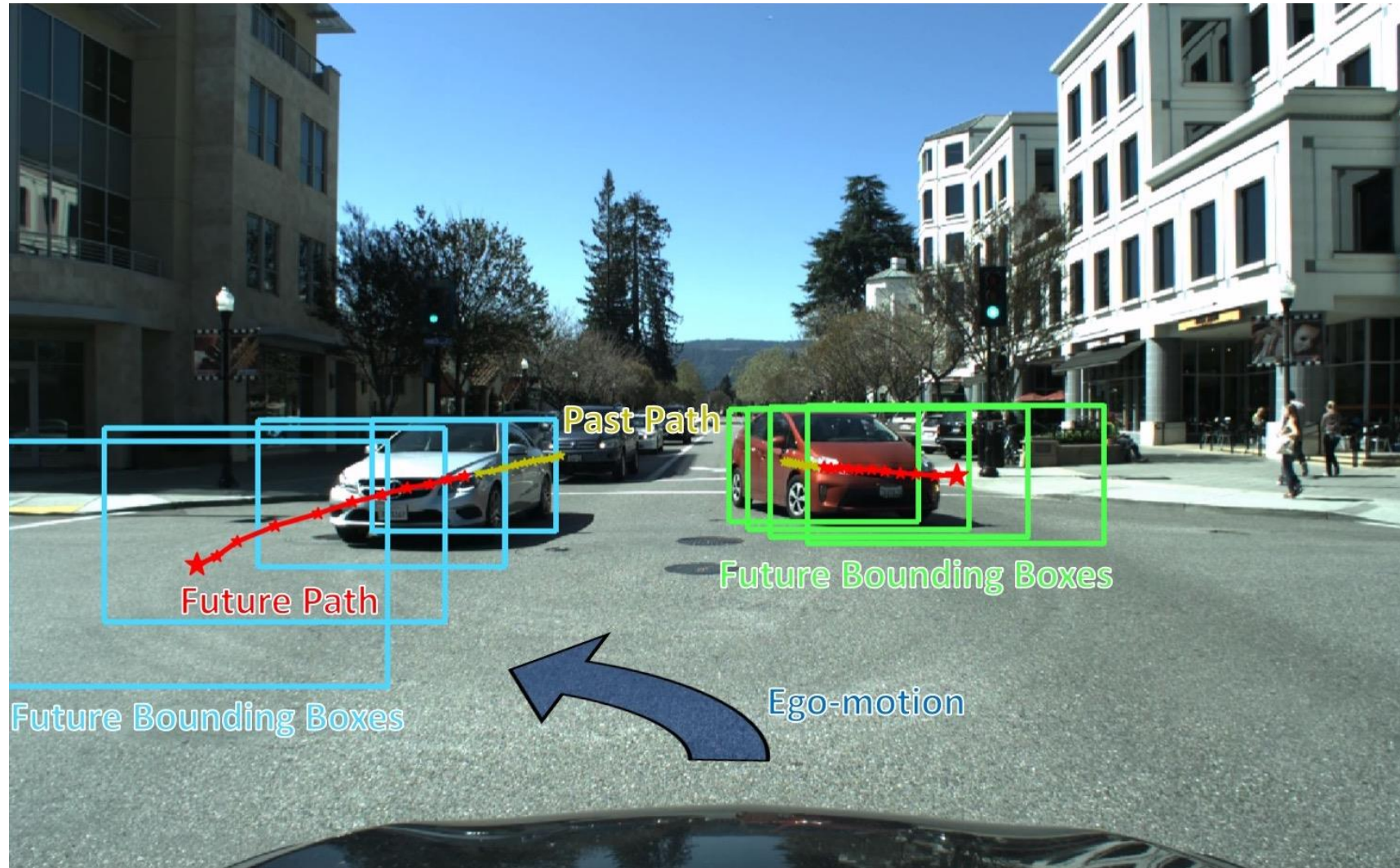
Autonomous Vehicles

- Detection & Segmentation
- Tracking



Autonomous Vehicles

- Future Prediction



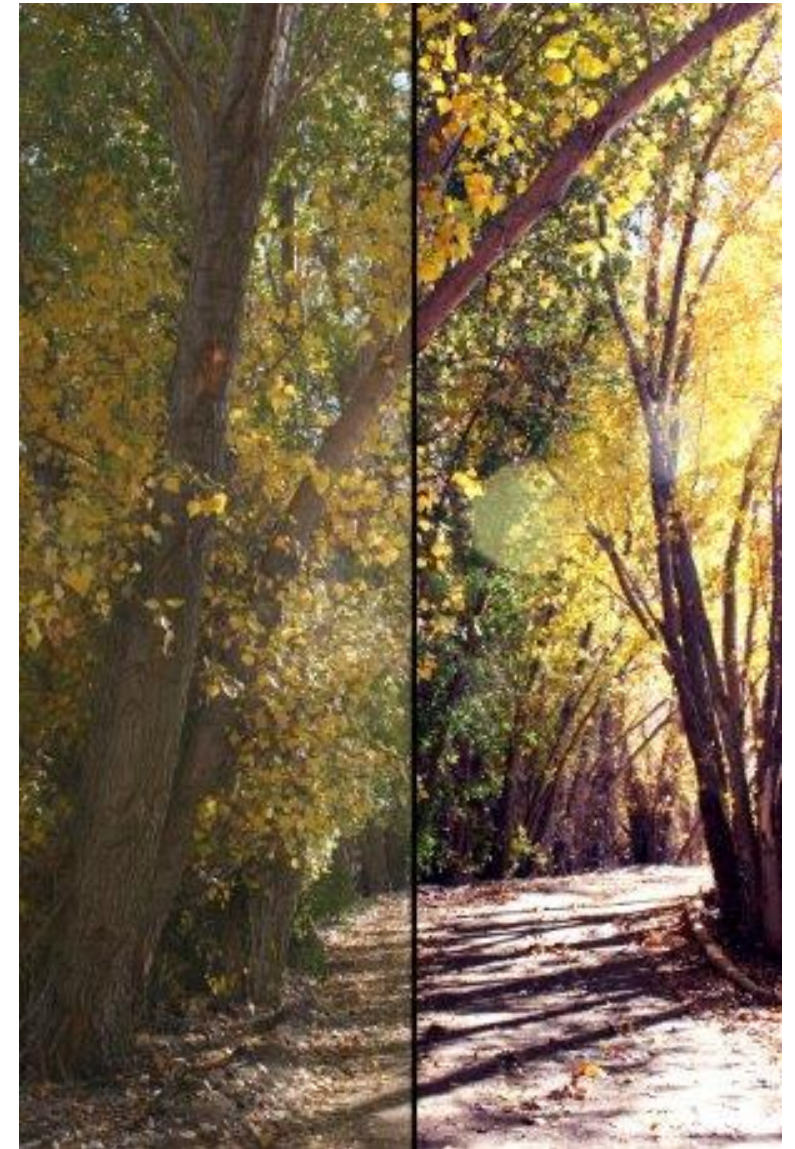
Autonomous Vehicles

- Gaze Estimation
- Mood Recognition
- Fatigue Monitoring



Photography

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



Special Effects

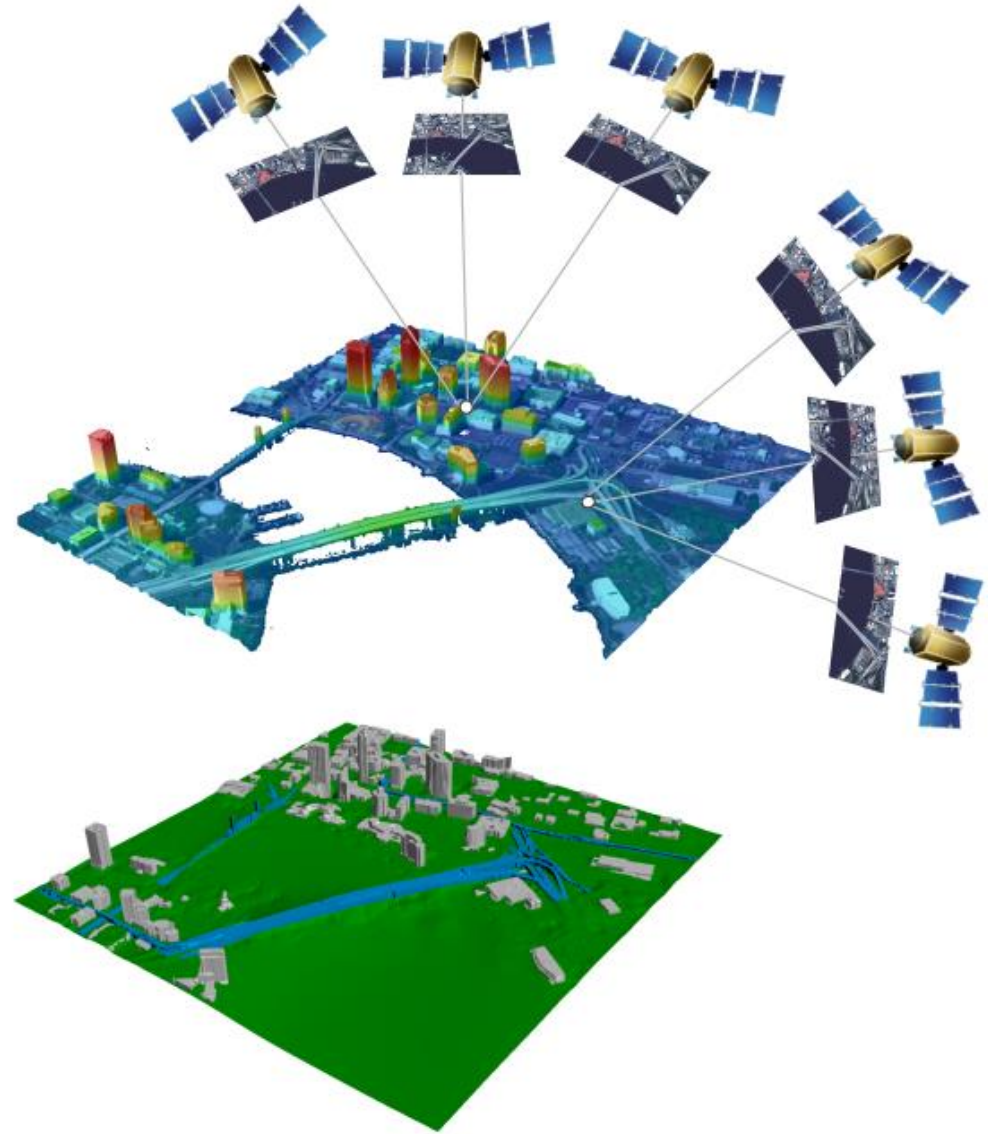
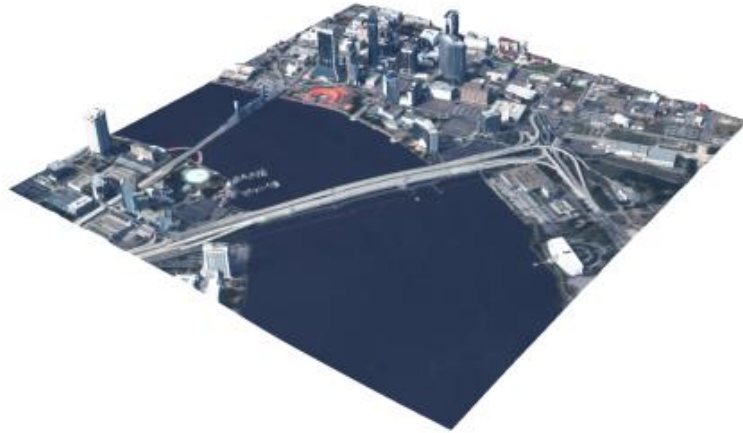
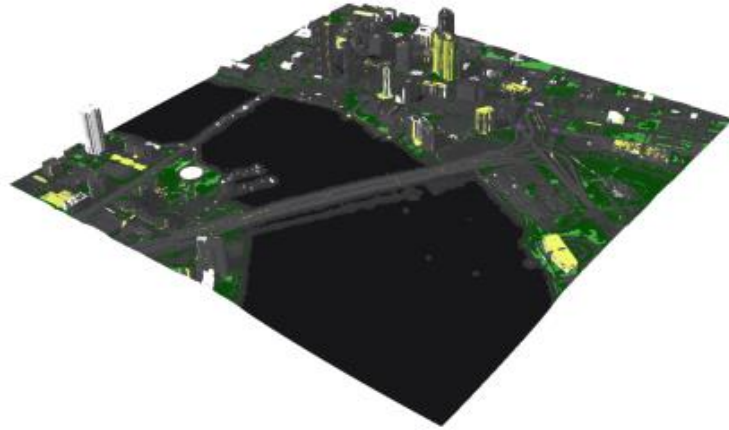


Special Effects (3D)

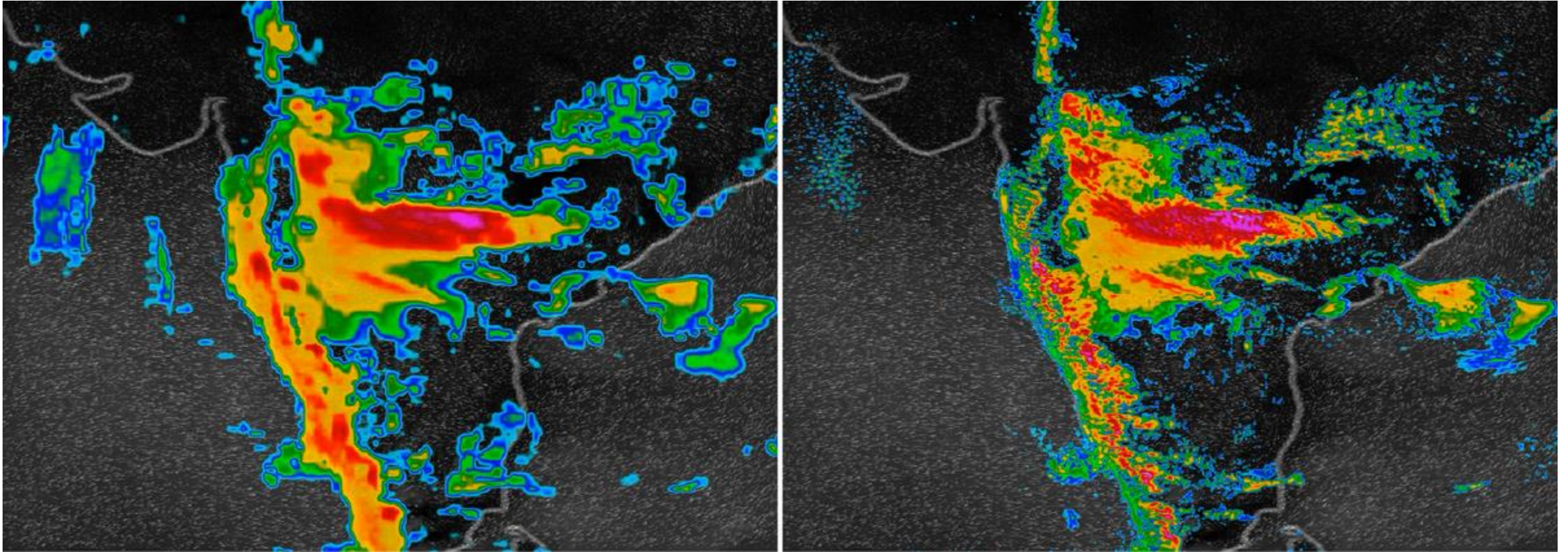


Map & 3D Model Generation

- 3D reconstruction
- Structure from motion

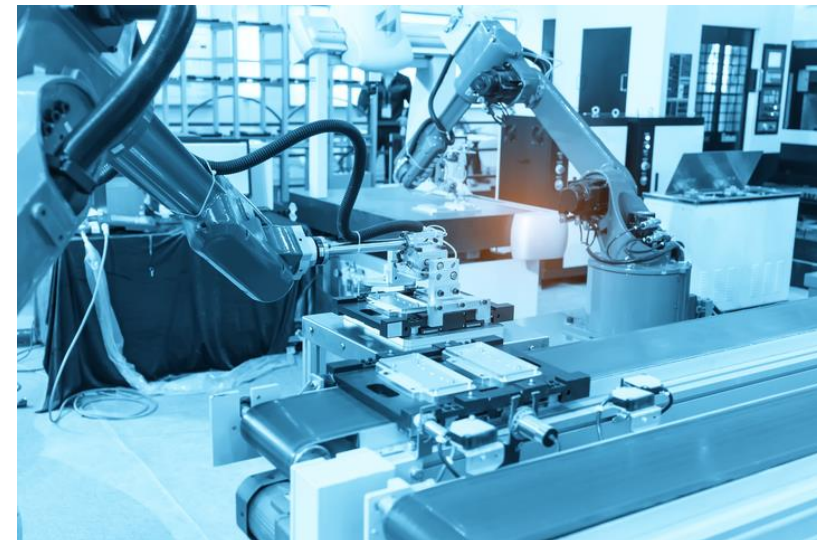


Weather Forecasting



- Spectral image analysis
- Image segmentation

Robotics

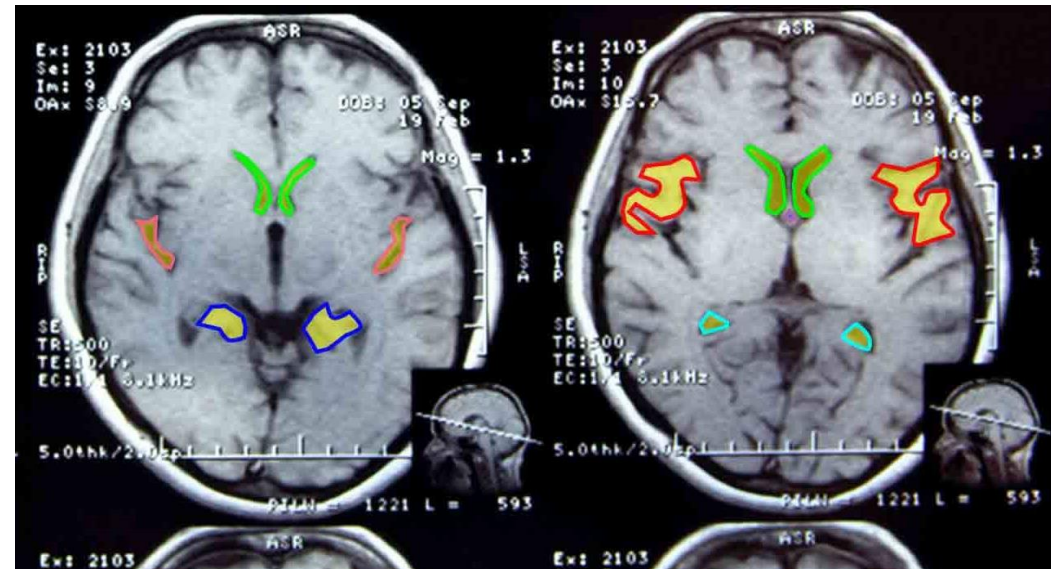


Quality Control & Monitoring



Healthcare

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



Carnegie Mellon University

FPS: 24

Confidence: 92%

Behavior Indicators

- Depression
- Smile Dynamics
- Gaze avoidance
- Post-Traumatic
- Negative expressions (male)
- Negative expressions (female)
- Head rotation variation
- Schizophrenia
- Expressivity
- Suicidality
- Pitch variability

MultiSense
Powered by OpenFace
Powered by COVAREP

The interface displays a video feed of a person's face with a green bounding box. On the left, there are several line graphs for 'Lips', 'Brows', 'Other', 'Head pose', 'Eye gaze', 'Audio', and 'Pitch'. On the right, there are horizontal bars for various behavioral indicators. The bottom right corner shows a heart icon and '126 bpm'.

2014-03-14 21:50:43

breath monitor

126 bpm

The interface shows a grayscale video feed of a person's face. A purple waveform is overlaid on the right side, labeled 'breath monitor'. A red heart icon and '126 bpm' are displayed in the bottom left corner. A timestamp '2014-03-14 21:50:43' is shown in a red box at the top left.

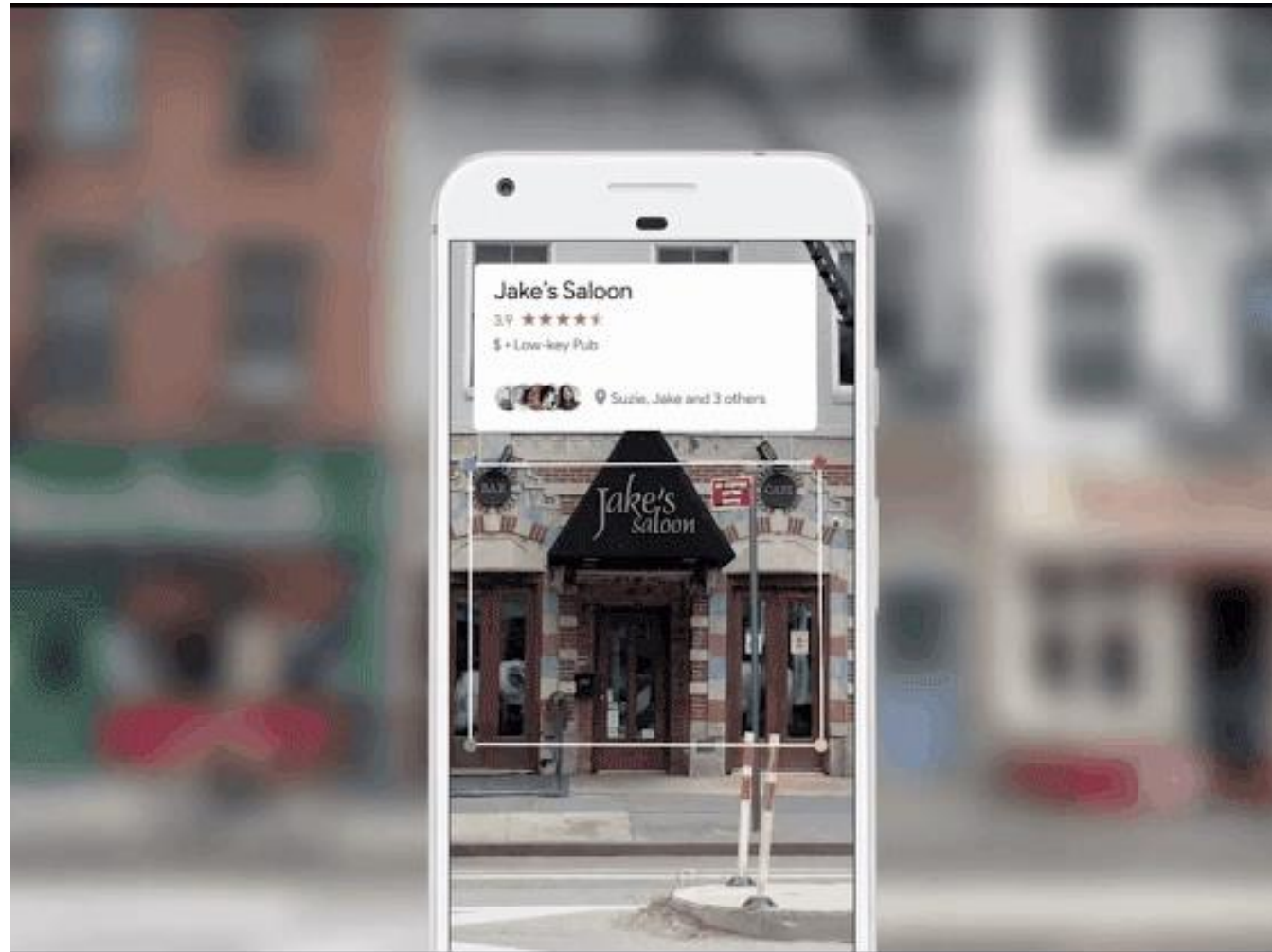
Neuromarketing

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



Visual Search

- <https://lens.google.com/>



Gaming / Human Computer Interaction

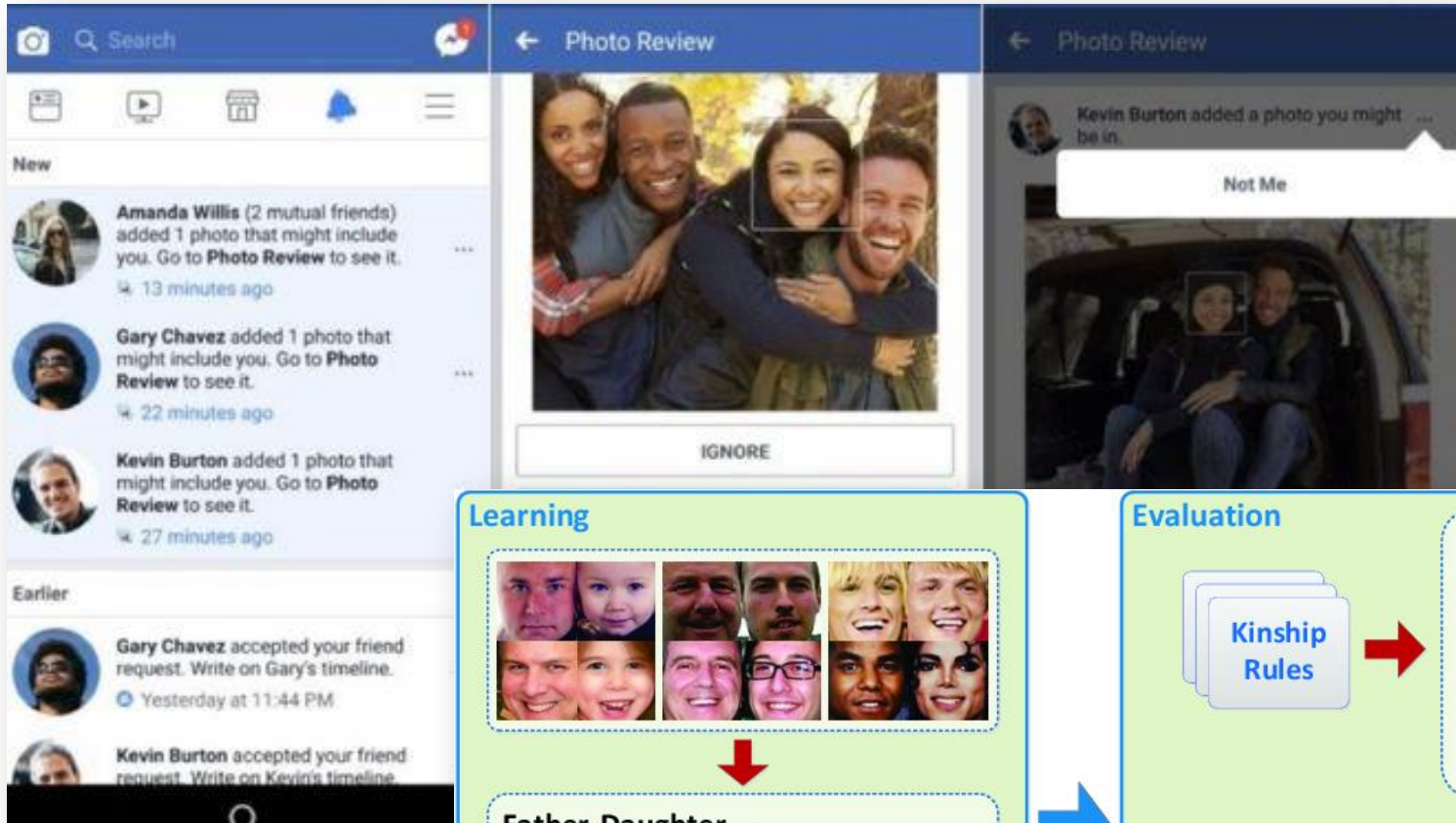


- Pose and motion tracking
- Gesture recognition

Sports Analytics



Image Analytics: Identity Recognition / Tagging



- Face detection
- Face recognition
- Kinship recognition

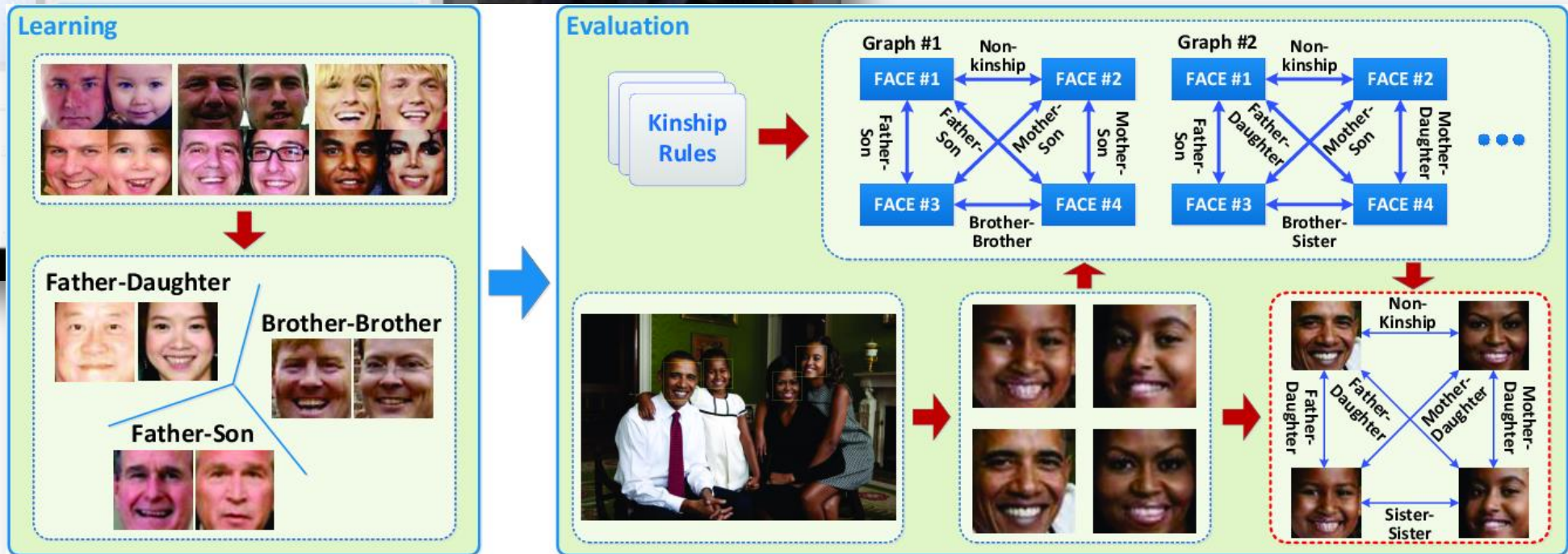


Image Analytics: Captioning



a yellow plate topped with meat and broccoli.



a zebra standing next to a zebra in a dirt field.



a stainless steel oven in a kitchen with wood cabinets.



two birds sitting on top of a tree branch.



an elephant standing next to rock wall.



a man riding a bike down a road next to a body of water.

Image Synthesis/Modification

0-18

19-29

30-39

40-49

50-59

60+



Style Transfer



+



=



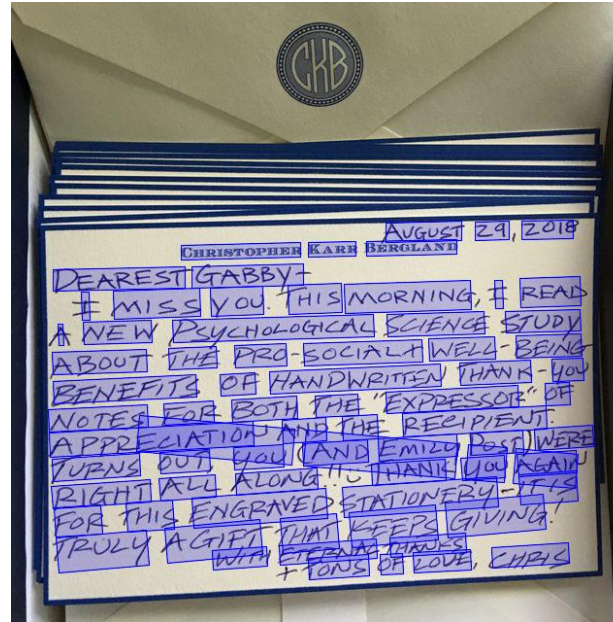
+



=

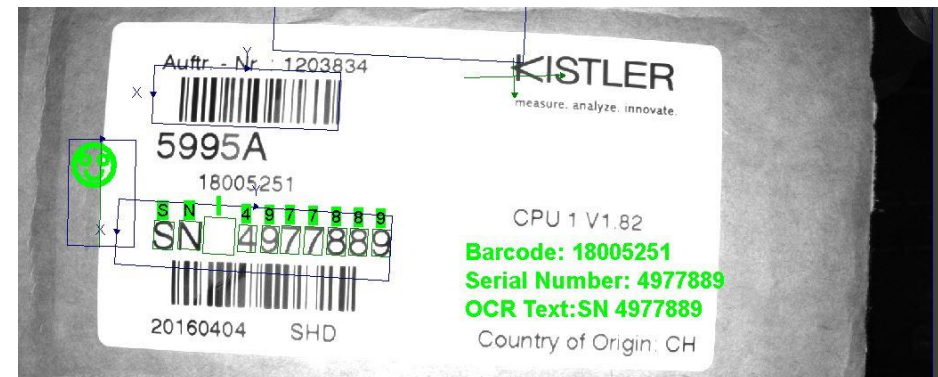


Text Recognition / Document Digitization



AUGUST 29 2018
CHRISTOPHER KARR BERGLAND

DEAREST GABBY I MISS you THIS MORNING I READ
A NEW PSYCHOLOGICAL SCIENCE STUDY
ABOUT THE PRO SOCIALE WELL BEING
BENEFITS OF HANDWRITTEN THANK You
NOTES ECATIONS FOR BOTH THE EXPRESSOR OF
APPRECIATION AND THE RECIPIENT
TURNS OUT you CAND DEMONIZ POST WERE
RIGHT ALL ALONG THANK YOU AGAIN
FOR THIS ENGRAVED STATIONERY ITIS
TRULY AGIFT THAT KEEPS GIVING
WITH ETERNAL THANKS CHes



And many more applications...

