Computer Vision (CSI4116)

Introduction

Lecture 01

https://www.youtube.com/watch?v=eQLcDmfmGB0

Today

- Course overview
- Course requirements
- Introduction to computer vision

Course Overview (1)

- Instructor:
  Seon Joo Kim (seonjookim@yonsei.ac.kr)
- Office hours:
  By appointment, RM 723 (Eng4)
- Textbooks (reference):
  - Richard Szeliski,
    Computer Vision: Algorithms and Applications
    (draft available online)
  - Forsyth & Ponce, *Computer Vision: A Modern Approach*
  - Hartley & Zisserman, Multiview Geometry
- Class webpage: LearnUs

Course Overview (2)

- Grading

  - 3 projects : 40%
  - Midterm/Final Exams : 60%
  - Attendance
    * -1% per missed class.
      (-2% for 2hour classes)

* All projects will involve implementation (python).

Collaboration policy

- Feel free to discuss assignments with each other, but coding must be done individually
- Feel free to incorporate code or tips you find on the Web, provided this doesn't make the assignment trivial and you explicitly acknowledge your sources
- Remember: I can Google too .

I can use ChatGPT too

Course Requirements

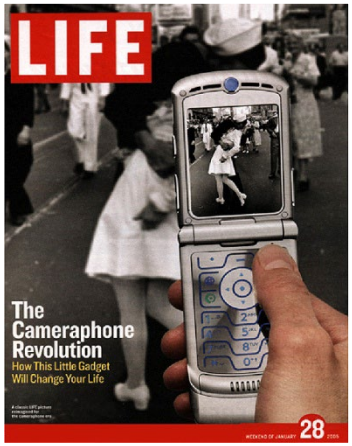
- Mathematical Background
  - Linear Algebra

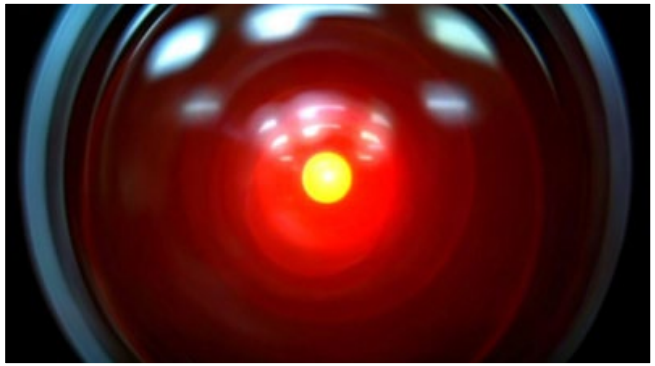
  (http://cseweb.ucsd.edu/classes/wi05/cse252a/linear_algebra_review.pdf)
  - Probability

  (http://cseweb.ucsd.edu/classes/wi05/cse252a/random_var_review.pdf)

- Programming Skills
  - python

Computer Vision





The goal of computer vision

- To bridge the gap between pixels and “meaning”

What we see

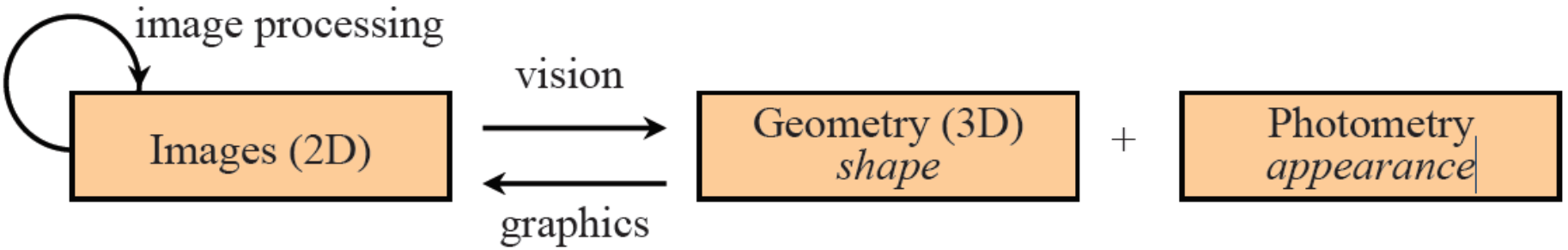
0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0

What a computer sees

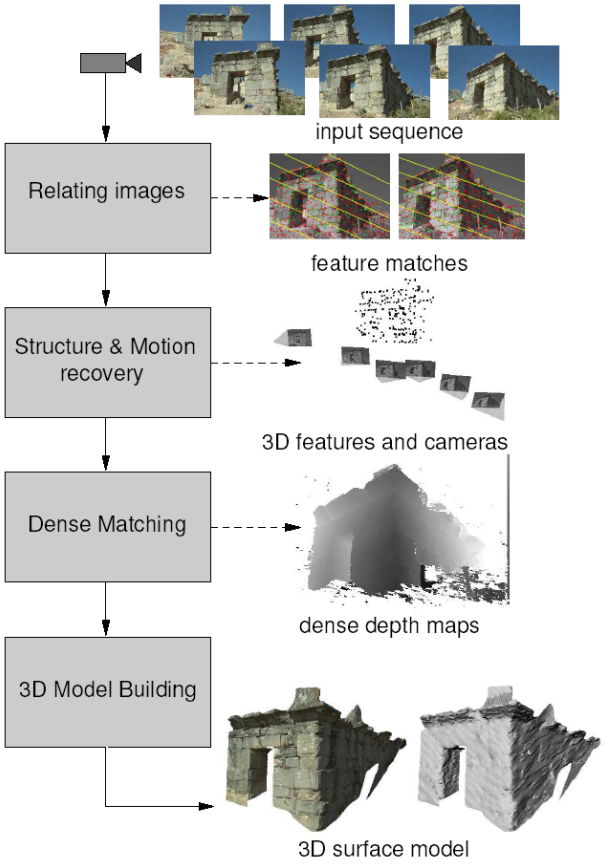
Source: S. Narasimhan

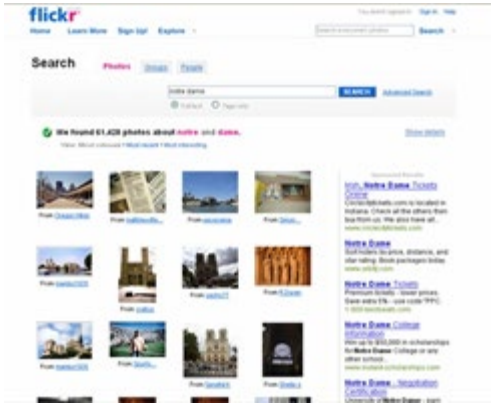
What kind of information can we extract from an image?

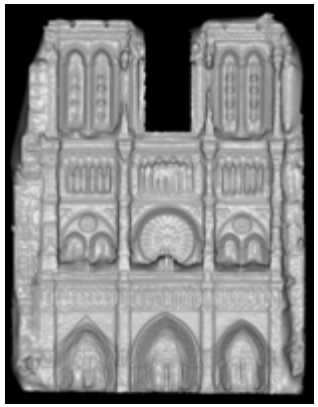
- Metric 3D information
- Semantic information



Vision as measurement device







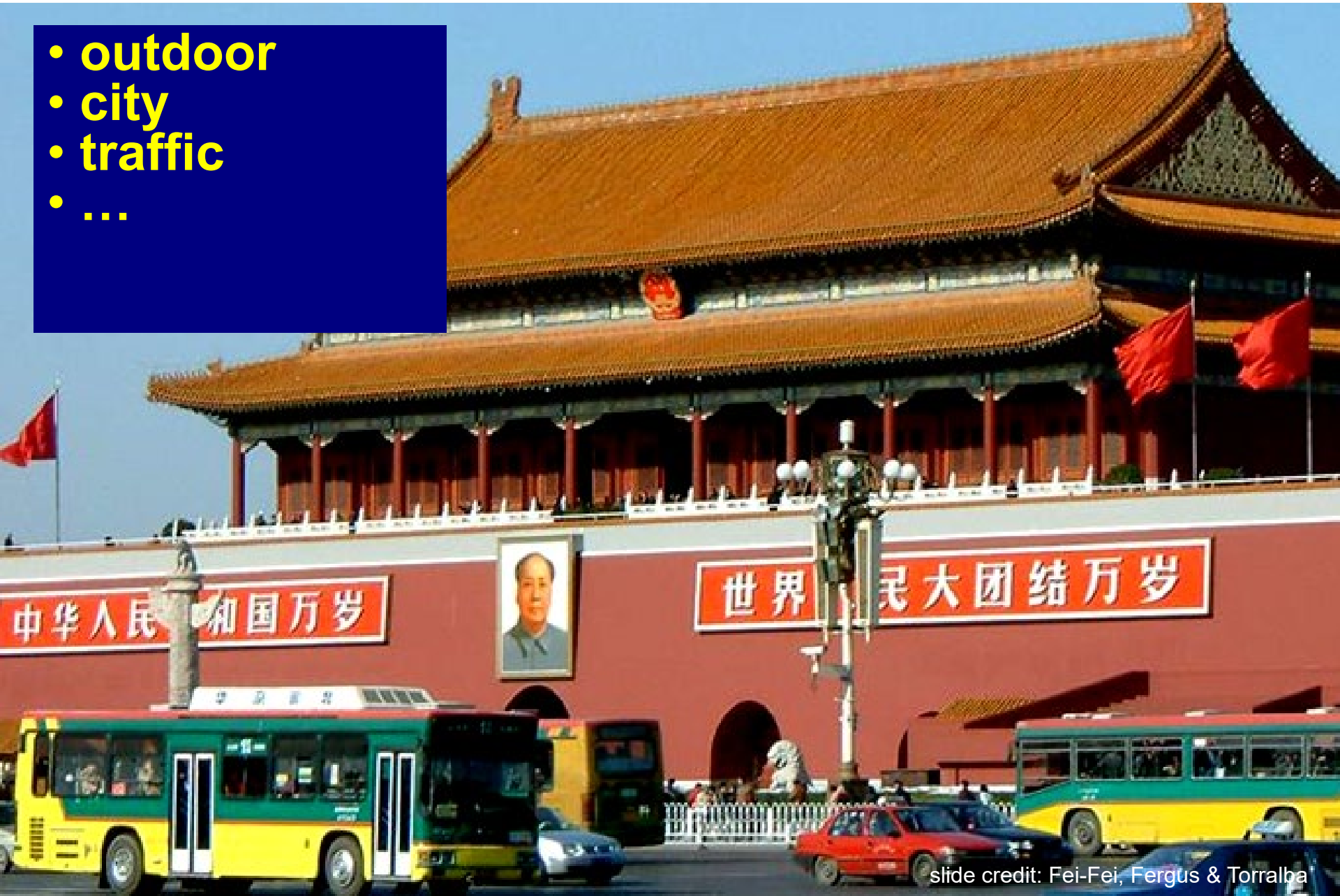
Vision as a source of semantic information

slide credit: Fei-Fei, Fergus & Torralba

Object categorization

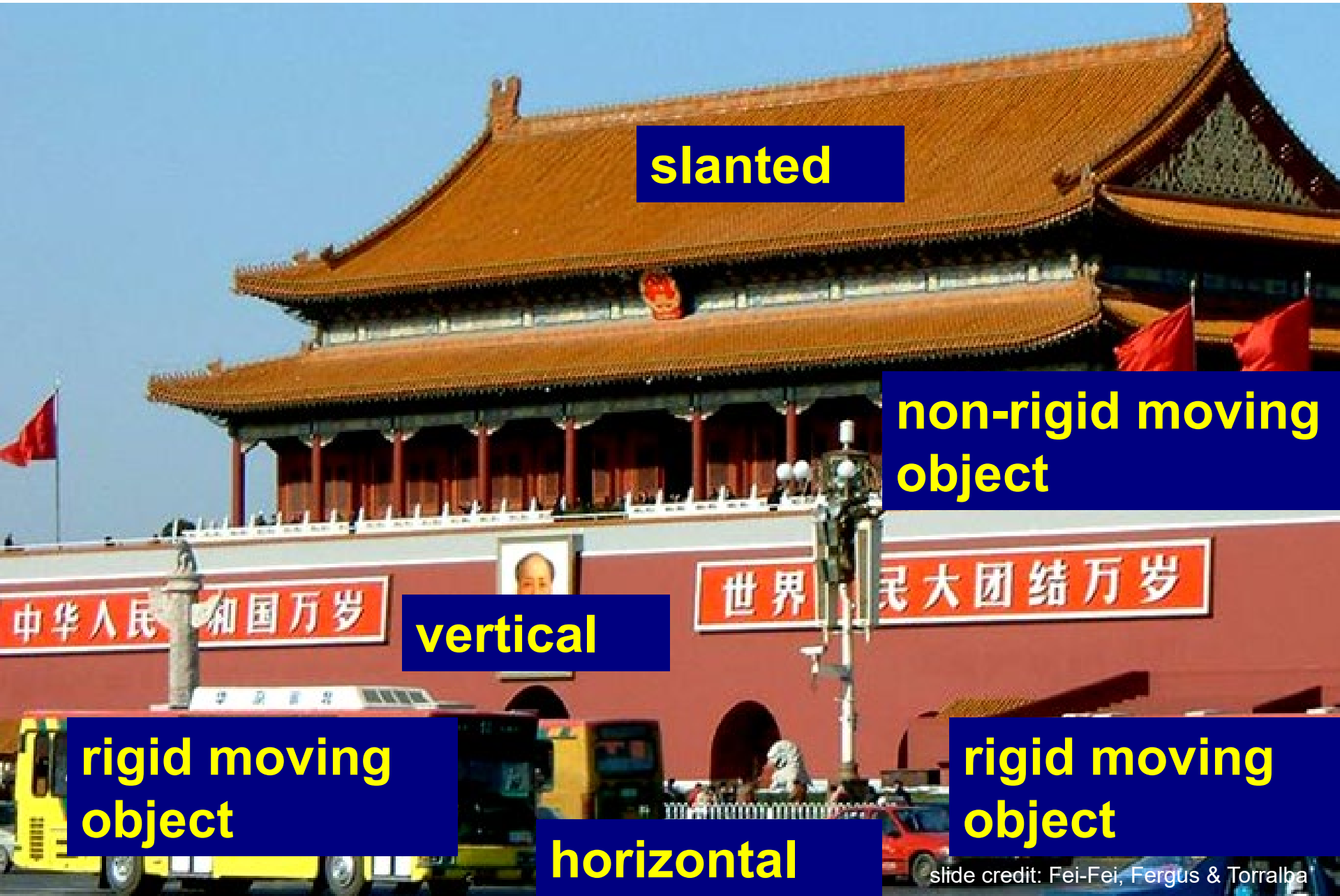
Scene and context categorization

- outdoor
- city
- traffic
- …



slide credit: Fei-Fei, Fergus & Torralba

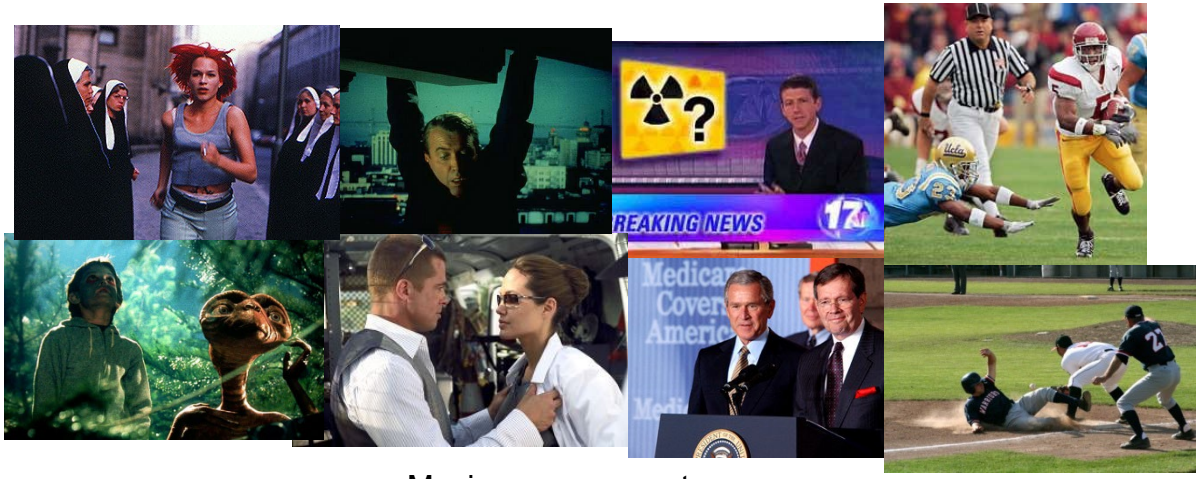
Qualitative spatial information



Why study computer vision?

- Vision is useful: Images and video are everywhere!

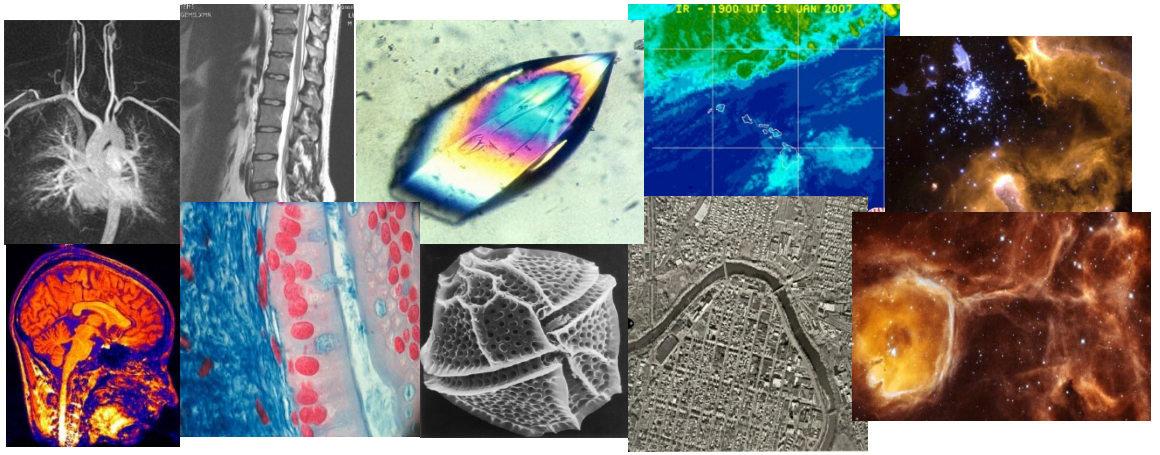
Personal photo albums



Movies, news, sports

Google Image Search Picasa™ flickr webshots picsearch™ YouTube Broadcast Yourself™

Surveillance and security

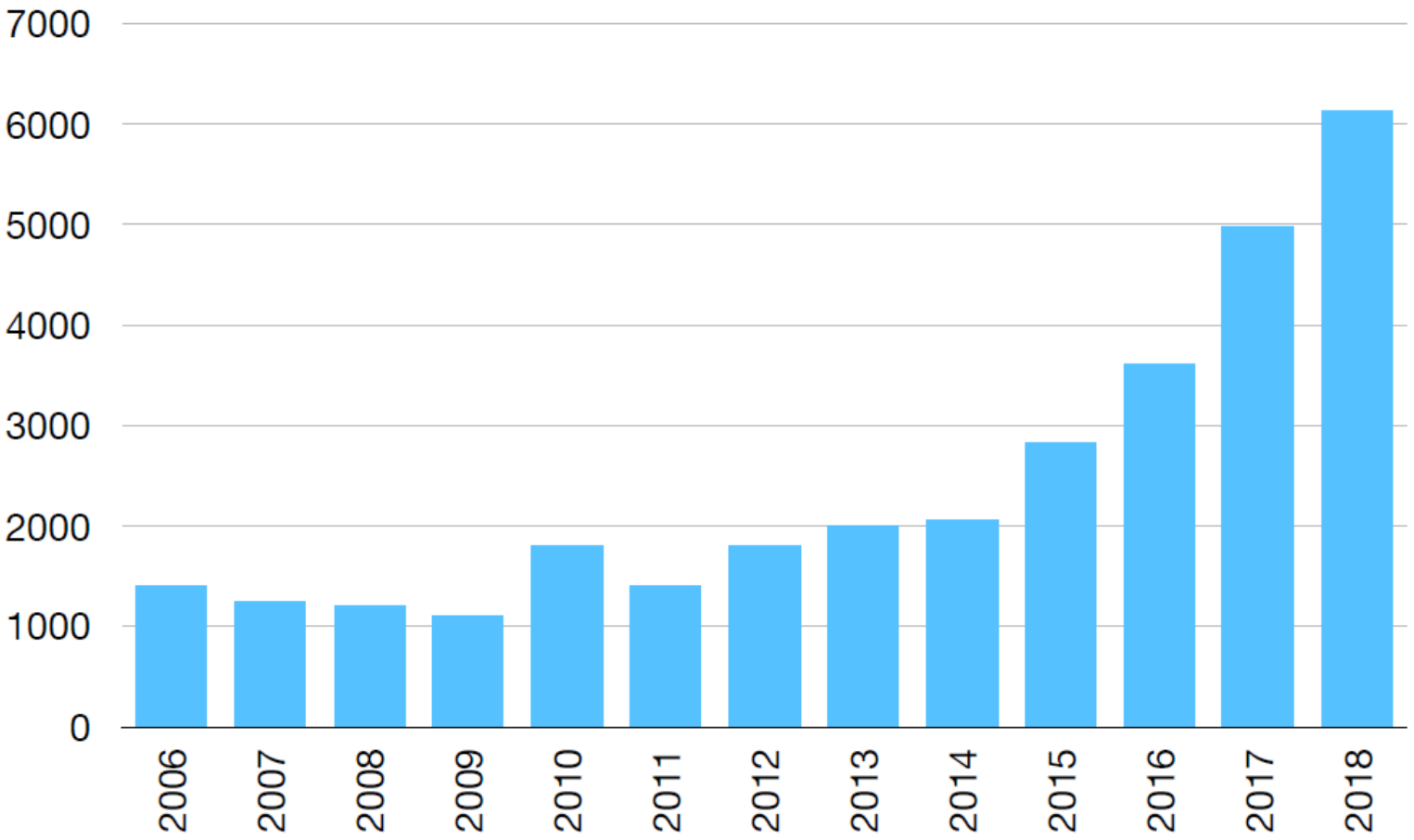


Medical and scientific images

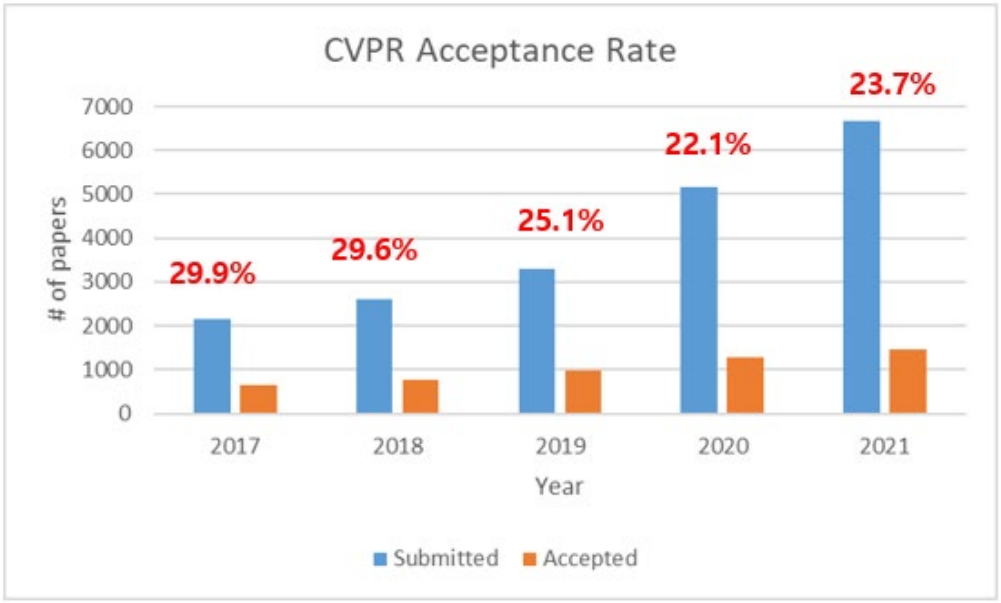
Why study computer vision?

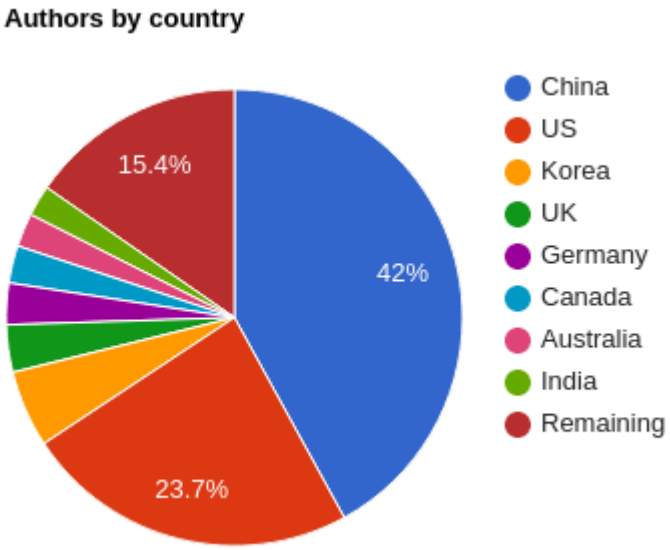
- Vision is useful
- Vision is interesting
- Vision is difficult
  - Half of primate cerebral cortex is devoted to visual processing
- Vision is hot! (mainly due to the advance in machine,…, well deep learning techniques)

CVPR Attendance

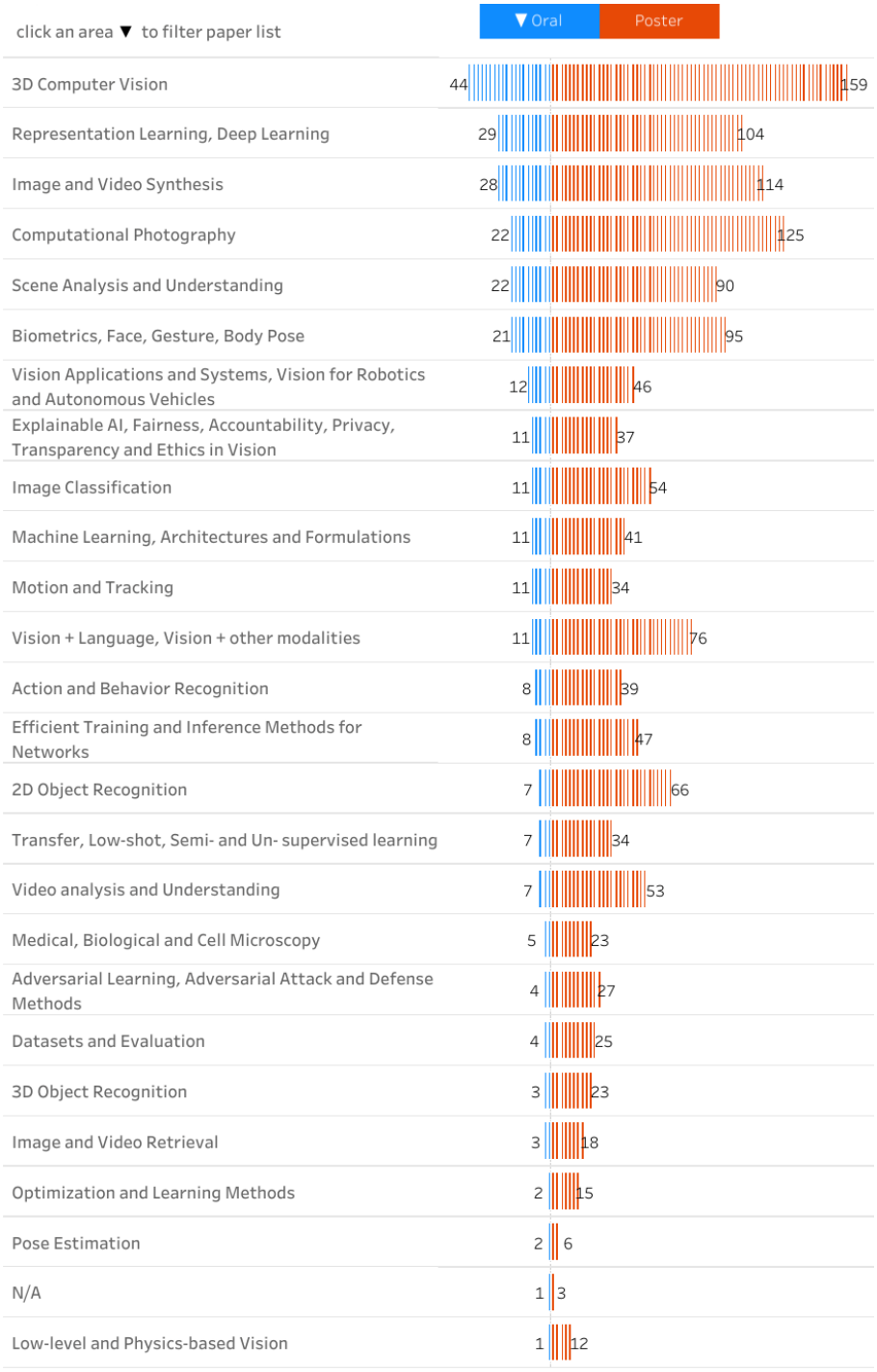


CVPR 2021 (1663 papers, 23.6%)



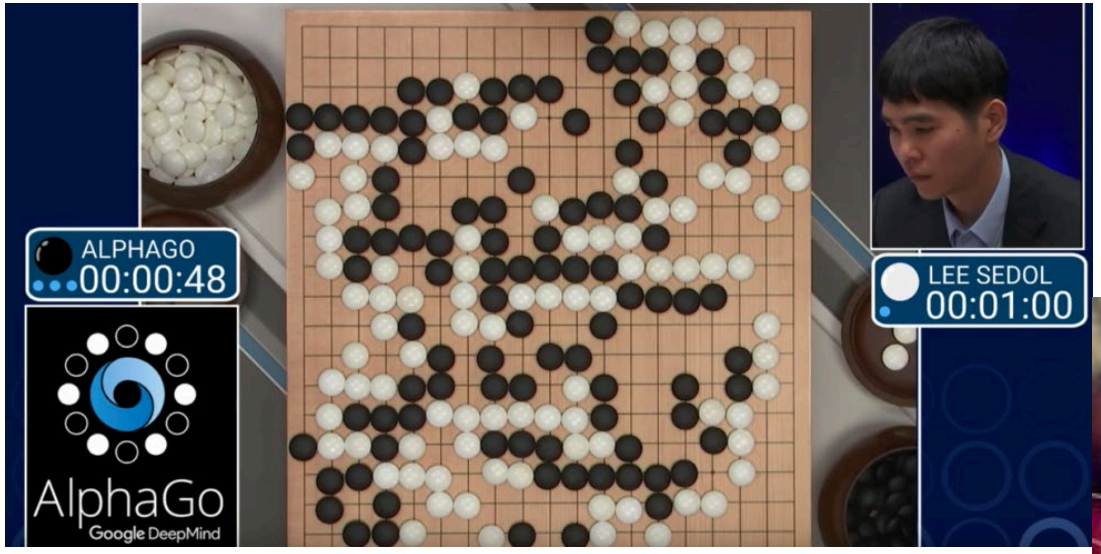


CVPR 2021

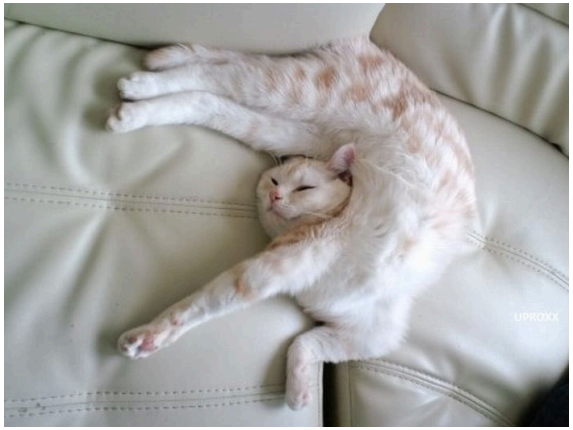

click an area ▼ to filter paper list
▼ Oral
Poster
3D Computer Vision 44 159
Representation Learning, Deep Learning 29 104
Image and Video Synthesis 28 114
Computational Photography 22 125
Scene Analysis and Understanding 22 90
Biometrics, Face, Gesture, Body Pose 21 95
Vision Applications and Systems, Vision for Robotics and Autonomous Vehicles 12 46
Explainable AI, Fairness, Accountability, Privacy, Transparency and Ethics in Vision 11 37
Image Classification 11 54
Machine Learning, Architectures and Formulations 11 41
Motion and Tracking 11 34
Vision + Language, Vision + other modalities 11 76
Action and Behavior Recognition 8 39
Efficient Training and Inference Methods for Networks 8 47
2D Object Recognition 7 66
Transfer, Low-shot, Semi- and Un- supervised learning 7 34
Video analysis and Understanding 7 53
Medical, Biological and Cell Microscopy 5 23
Adversarial Learning, Adversarial Attack and Defense Methods 4 27
Datasets and Evaluation 4 25
3D Object Recognition 3 23
Image and Video Retrieval 3 18
Optimization and Learning Methods 2 15
Pose Estimation 2 6
N/A 1 3
Low-level and Physics-based Vision 1 12

Why is computer vision difficult?

Alpha-Go vs. Image Recognition



Difficulty



Challenges: viewpoint variation

Michelangelo 1475-1564

slide credit: Fei-Fei, Fergus & Torralba

Challenges: illumination

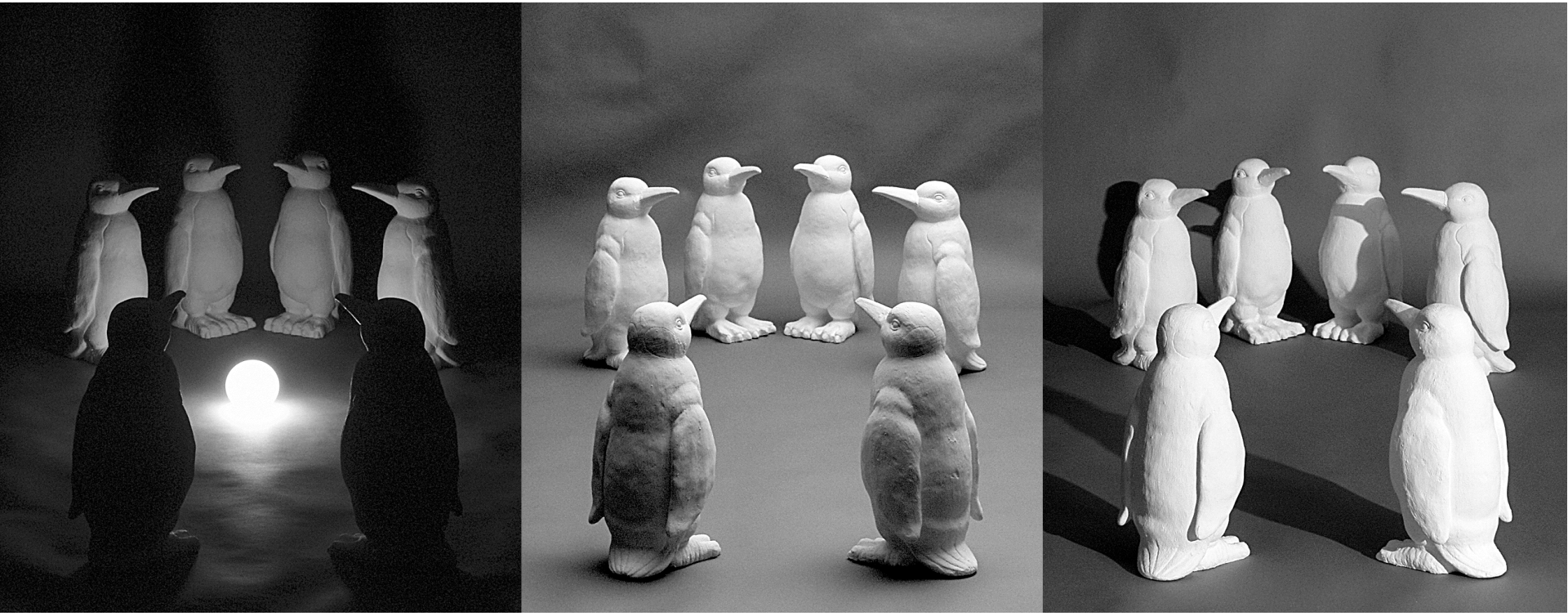
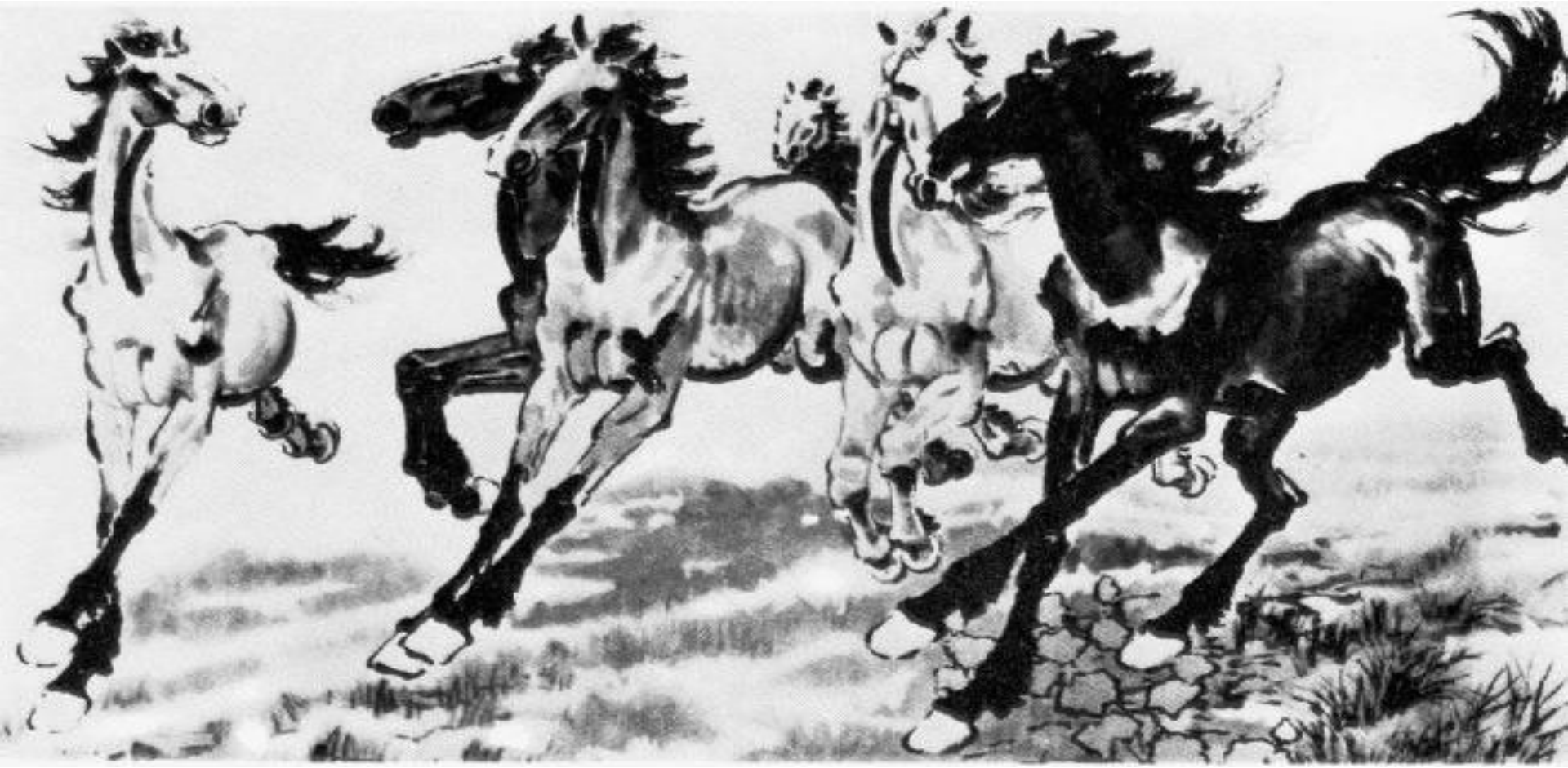


image credit: J. Koenderink

Challenges: scale

slide credit: Fei-Fei, Fergus & Torralba

Challenges: deformation



Xu, Beihong 1943

slide credit: Fei-Fei, Fergus & Torralba

Challenges: occlusion

Magritte, 1957

slide credit: Fei-Fei, Fergus & Torralba

Challenges: background clutter

Challenges: Motion

Challenges: object intra-class variation

slide credit: Fei-Fei, Fergus & Torralba

Challenges or opportunities?

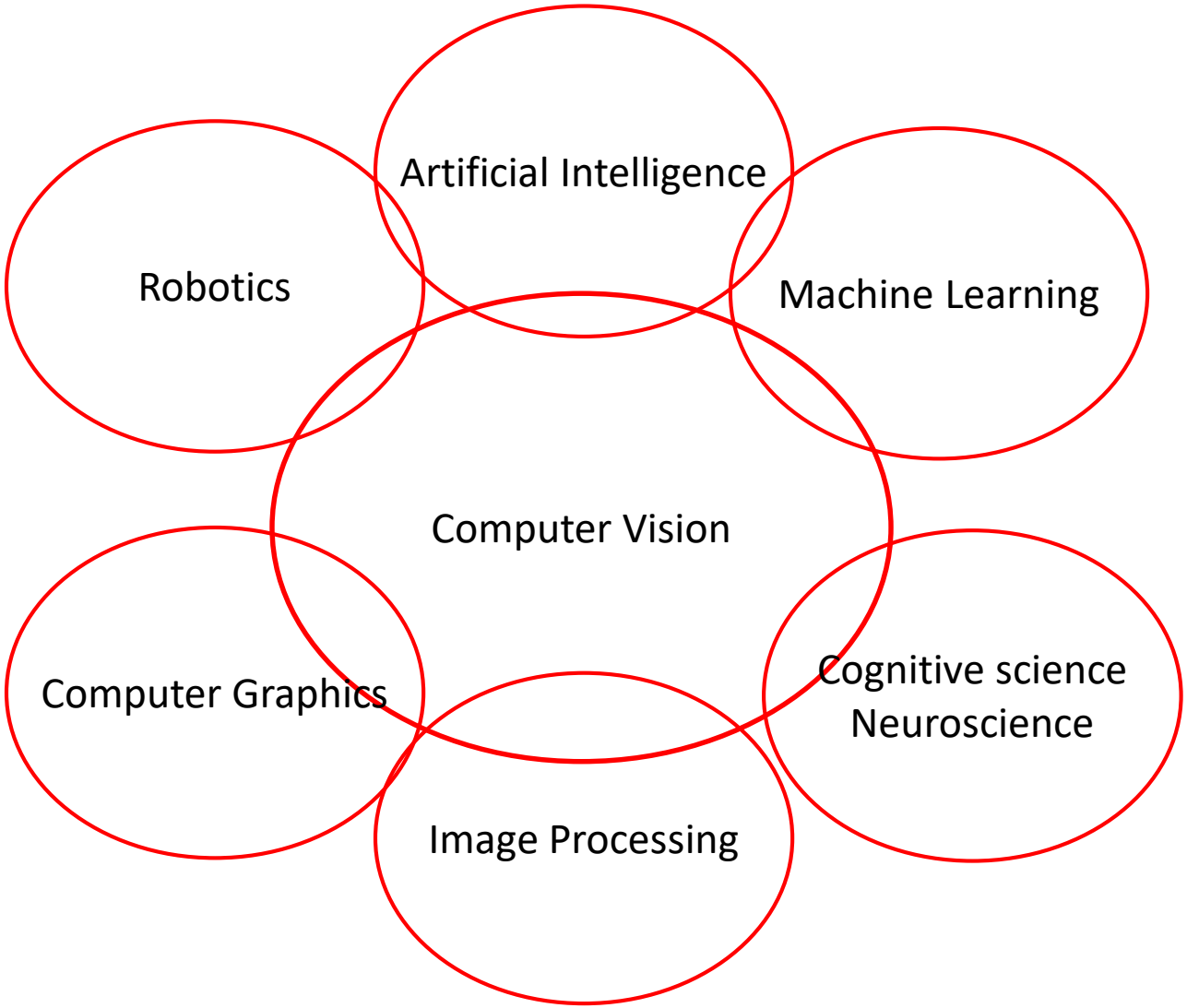
- Images are confusing, but they also reveal the structure of the world through numerous cues
- Our job is to interpret the cues!

Bottom line

- Perception is an inherently ambiguous problem
  - Many different 3D scenes could have given rise to a particular 2D picture

- Possible solutions
  - Bring in more constraints (more images)
  - Use prior knowledge about the structure of the world
- Need a combination of different methods

Connections to other disciplines

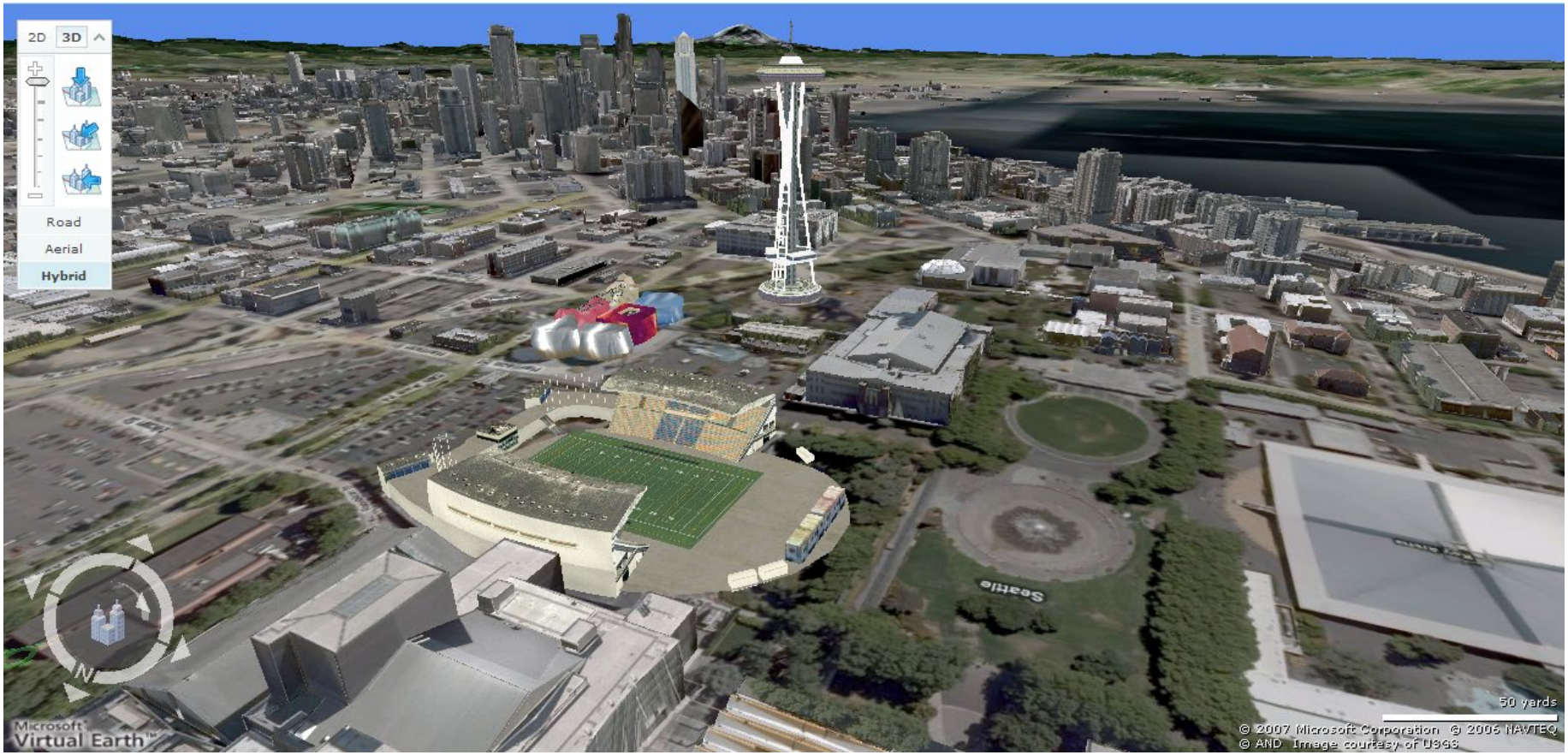


Computer Vision in the Real World

Special effects: shape and motion capture

Source: S. Seitz

3D urban modeling



Bing maps, Google Streetview

Source: S. Seitz

Face detection

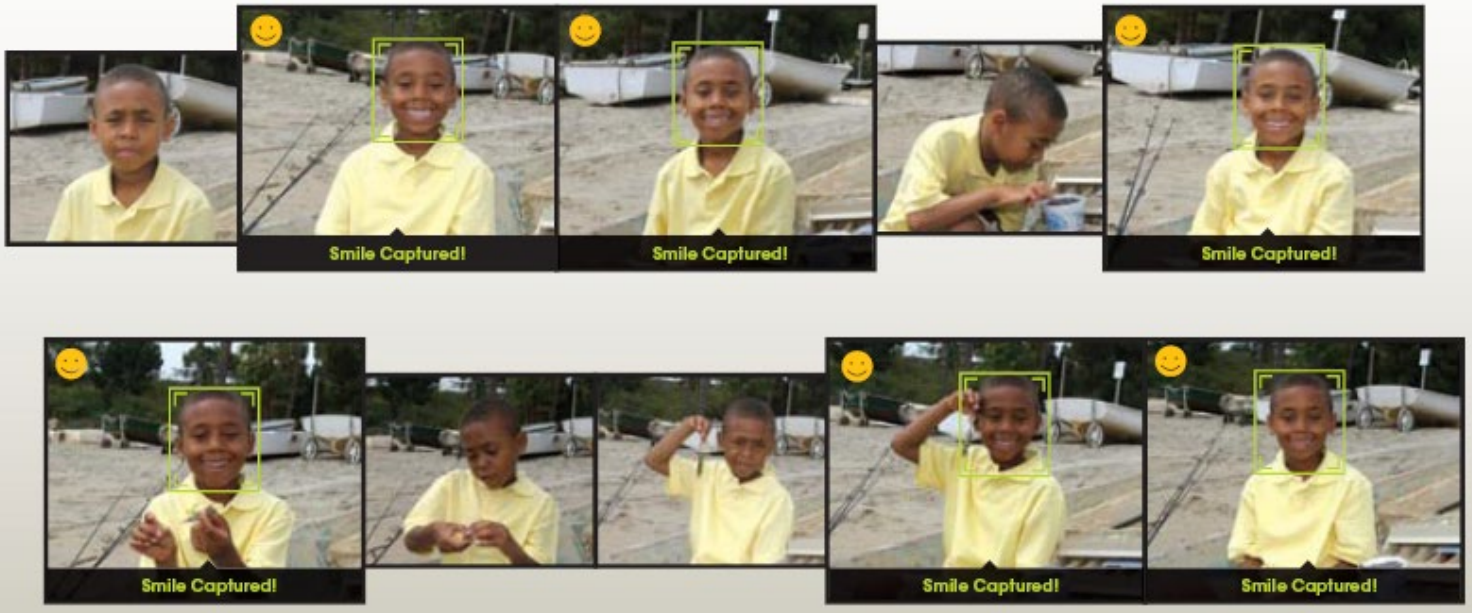
- Many new digital cameras now detect faces
  - Canon, Sony, Fuji, …

Source: S. Seitz

Smile detection

The Smile Shutter flow

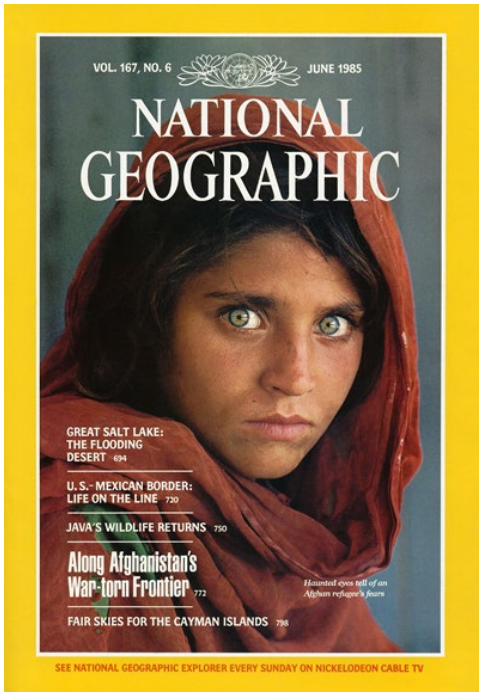
Imagine a camera smart enough to catch every smile! In Smile Shutter Mode, your Cyber-shot® camera can automatically trip the shutter at just the right instant to catch the perfect expression.



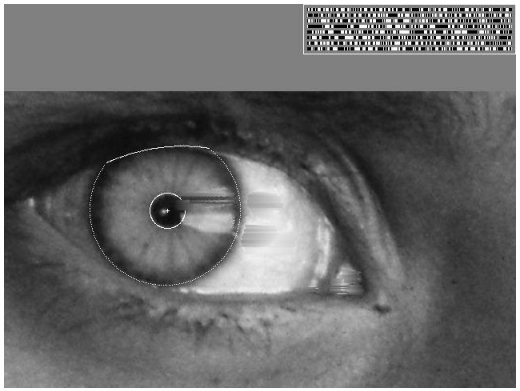
Sony Cyber-shot® T70 Digital Still Camera

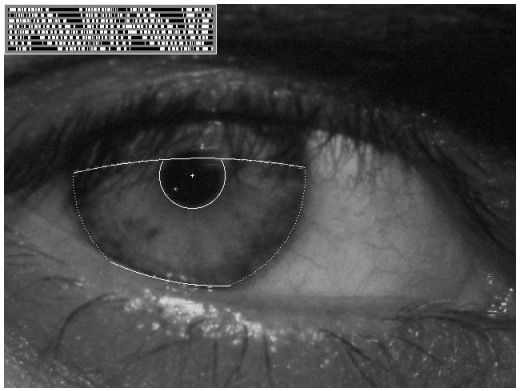
Source: S. Seitz

Biometrics



[How the Afghan Girl was Identified by Her Iris Patterns]





Source: S. Seitz

Biometrics

Fingerprint scanners on many new laptops, other devices

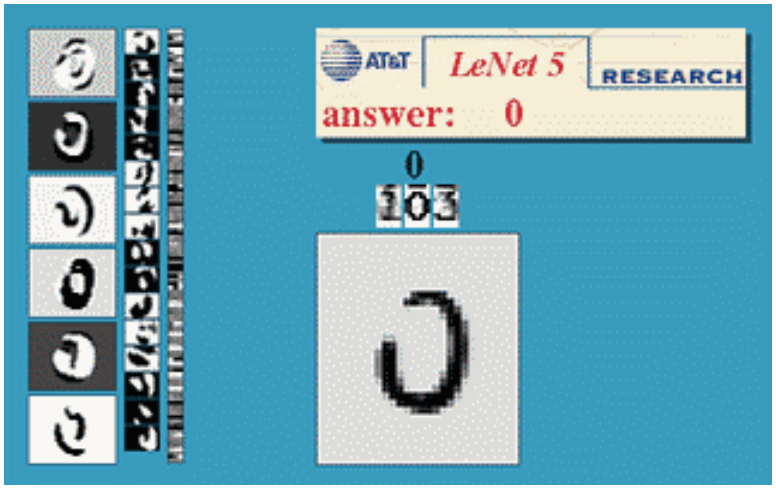
Face recognition systems now beginning to appear more widely
http://www.sensiblevision.com/

Source: S. Seitz

Optical character recognition (OCR)

Technology to convert scanned docs to text

- If you have a scanner, it probably came with OCR software



Digit recognition, AT&T labs

License plate readers
http://en.wikipedia.org/wiki/Automatic_number_plate_recognition

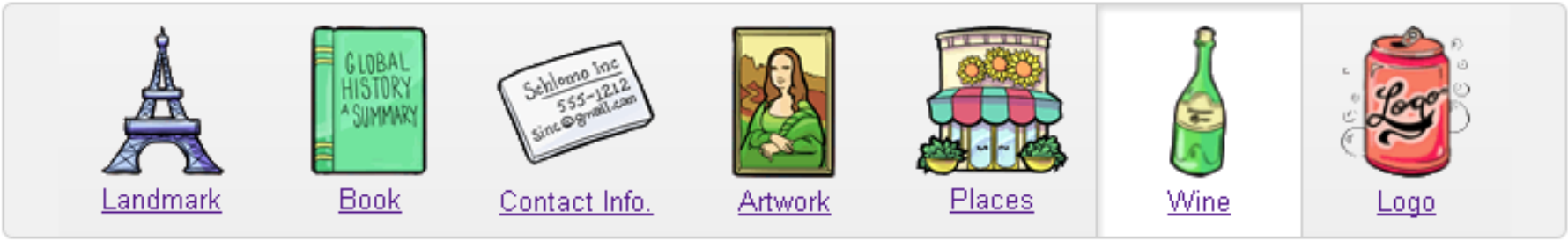
Source: S. Seitz

Mobile visual search: Google Goggles

Google Goggles in Action

Click the icons below to see the different ways Google Goggles can be used.

Google goggles labs



Automotive safety

- Mobileye: Vision systems in high-end BMW, GM, Volvo models
  - "In mid 2010 Mobileye will launch a world's first application of full emergency braking for collision mitigation for pedestrians where vision is the key technology for detecting pedestrians."

Source: A. Shashua, S. Seitz

Vision for robotics, space exploration

NASA'S Mars Exploration Rover Spirit captured this westward view from atop a low plateau where Spirit spent the closing months of 2007.

Vision systems (JPL) used for several tasks

- Panorama stitching
- 3D terrain modeling
- Obstacle detection, position tracking
- For more, read "Computer Vision on Mars" by Matthies et al.

Source: S. Seitz

Everything Changed with Deep Learning

Deep Learning

Computer Vision + Natural Language Processing

- image captioning

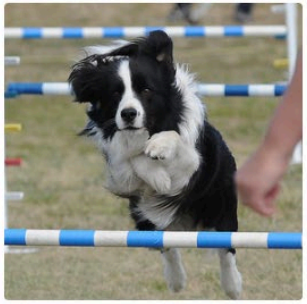
"man in black shirt is playing guitar."

"construction worker in orange safety vest is working on road."

"two young girls are playing with lego toy."

"boy is doing backflip on wakeboard."

"girl in pink dress is jumping in air."



"black and white dog jumps over bar."

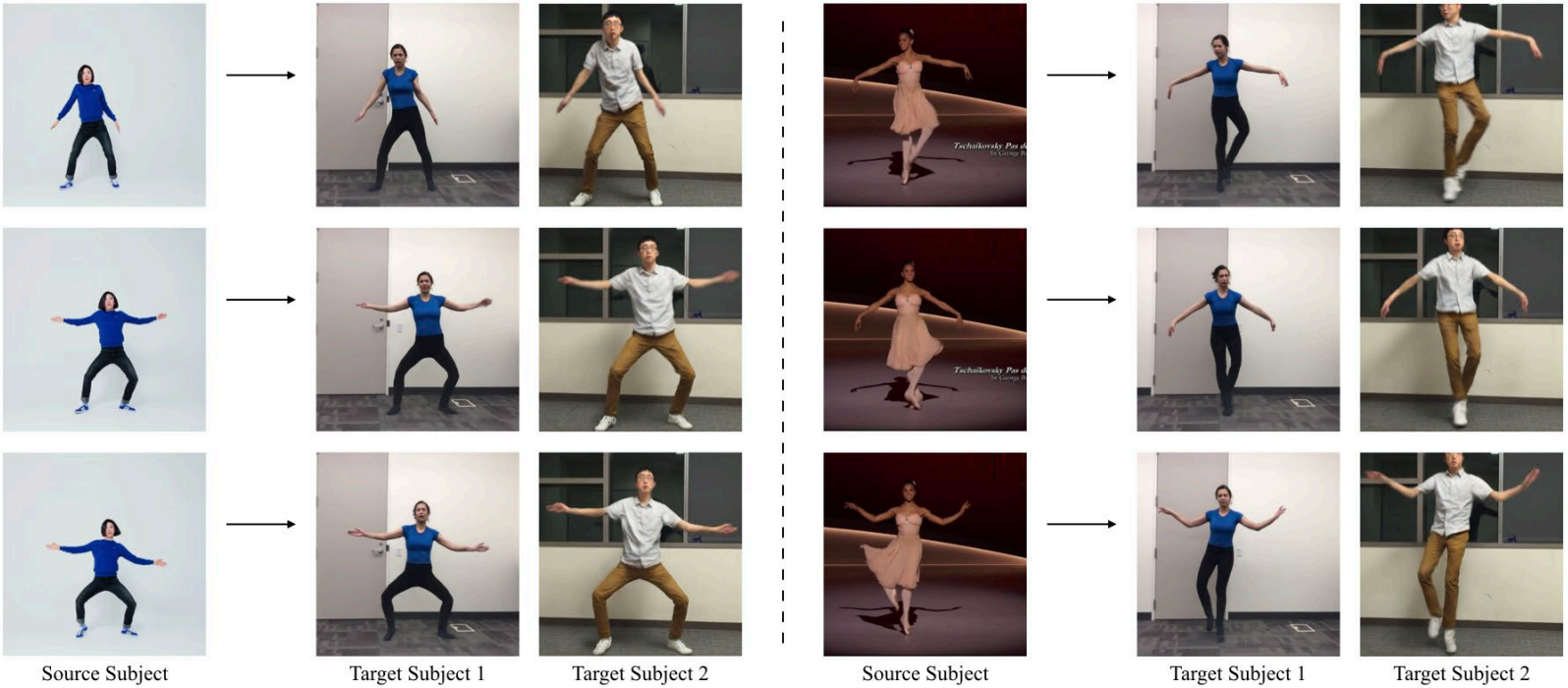
"young girl in pink shirt is swinging on swing."

"man in blue wetsuit is surfing on wave."

Amazon Go

https://www.youtube.com/watch?v=NrmMk1Myrxc

Everybody dance now

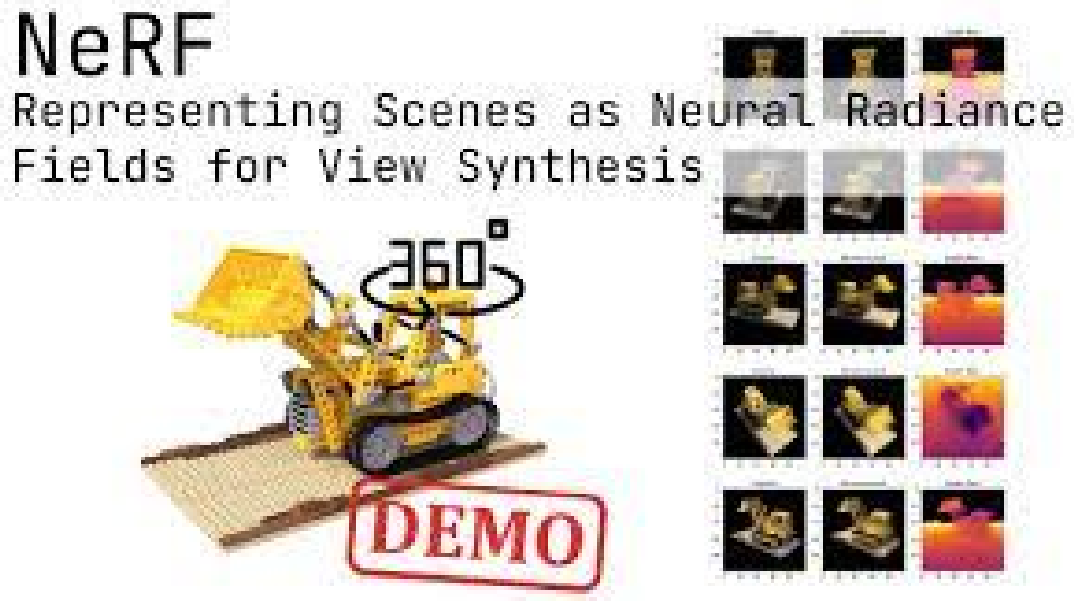


https://www.youtube.com/watch?time_continue=2&v=PCBTZh41Ris

DALL-E 2

DALL-E 2's interpretation of "A photo of an astronaut riding a horse."

NeRf



https://www.youtube.com/watch?v=DJ2hcC1orc4

Computer Vision, where to look for?

- Major Conferences :

  CVPR (Computer Vision and Pattern Recognition, every yr)

  ICCV (International Conference on Computer Vision, every 2 yr)

  ECCV (European Conference on Computer Vision, every 2 yr)
  NeurIPS (every year), ICML (every year), ICLR (every year)

  SIGGRAPH, SIGGRAPH Asia (computer graphics, every year)

- Major Journals:

  IEEE TPAMI (IEEE Transactions on Pattern Analysis and Machine Intelligence)

  IJCV (International Journal of Computer Vision)

  ACM TOG (Transactions on Graphics)

Excited???