ECE 5554/ ECE 4554
Computer Vision

Jia-Bin Huang

Electrical and Computer Engineering

Virginia Tech

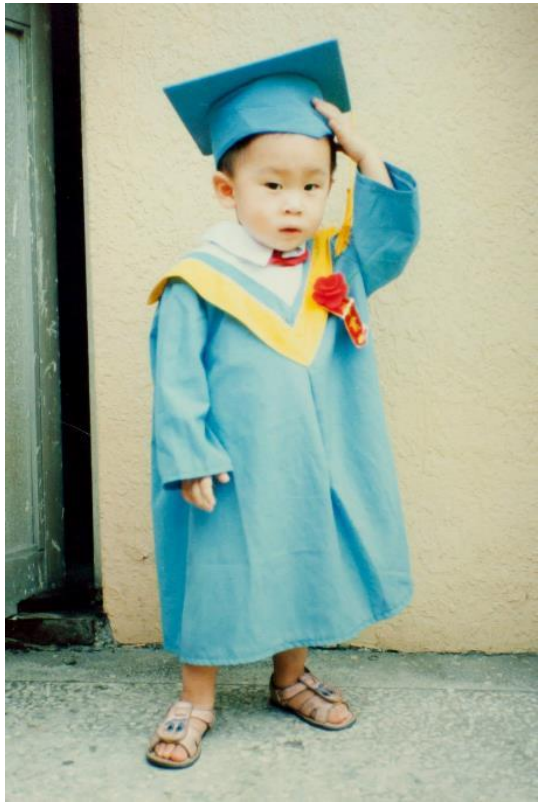
Today’s class

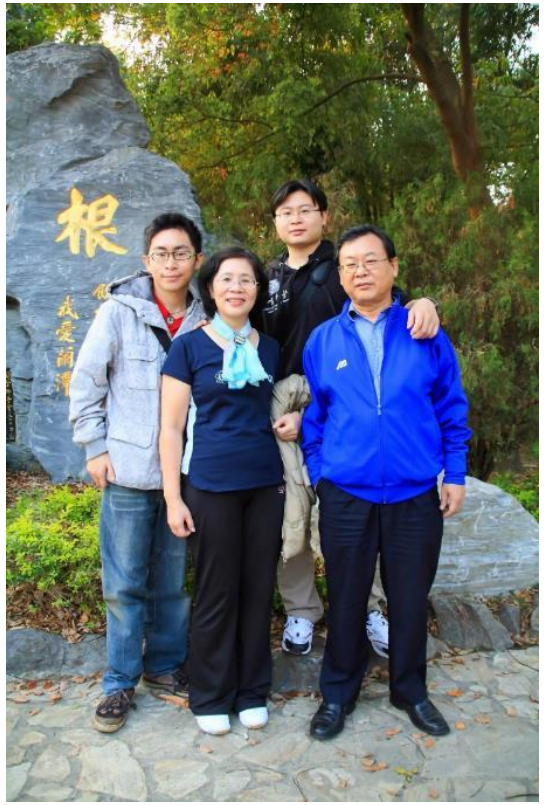
- A little about me
- A little about you
- Intro to computer vision
- Course logistics
- Questions

About me

- Born and raised in Taiwan

14 months old!





National Chiao-Tung University
B.S. in EE

Microsoft Research
Research Intern

Disney Research
Research Intern

UIUC
Ph.D. in ECE 2016

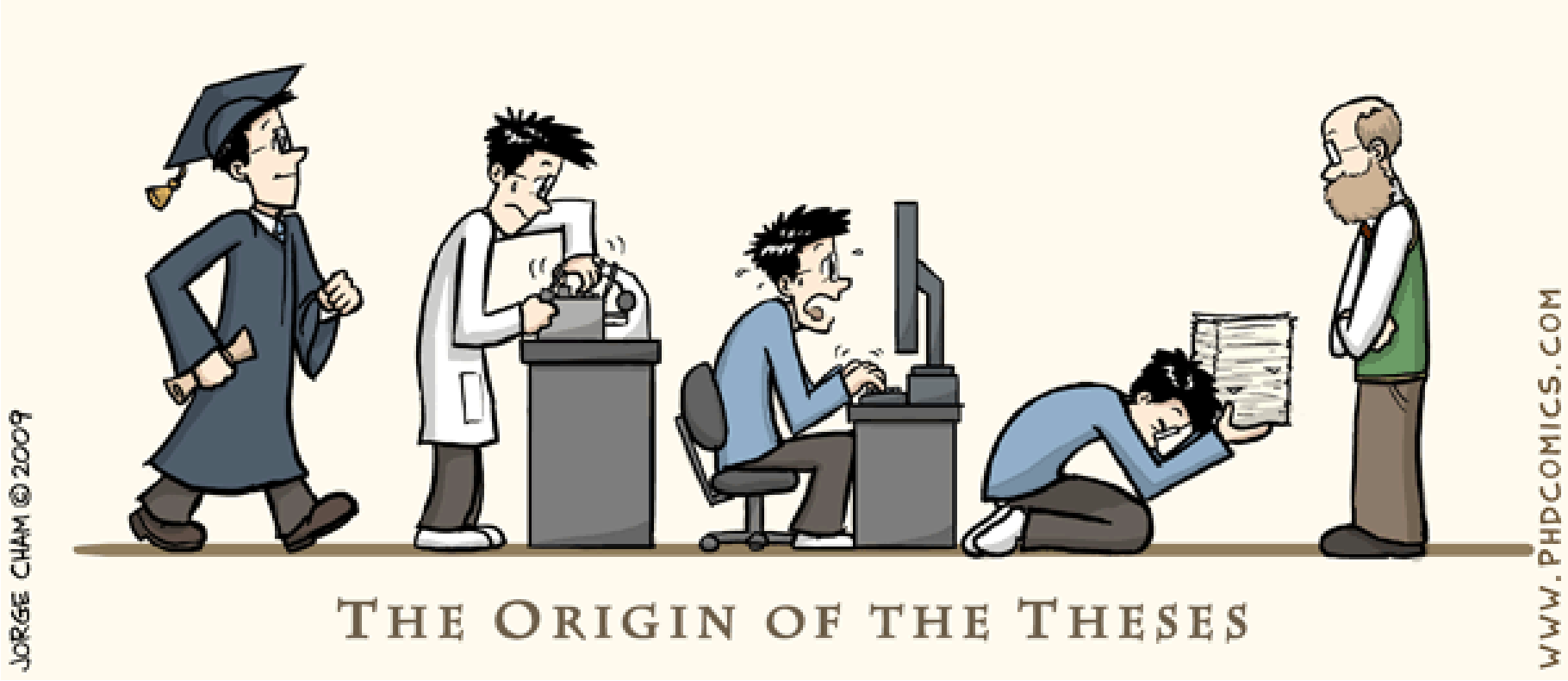
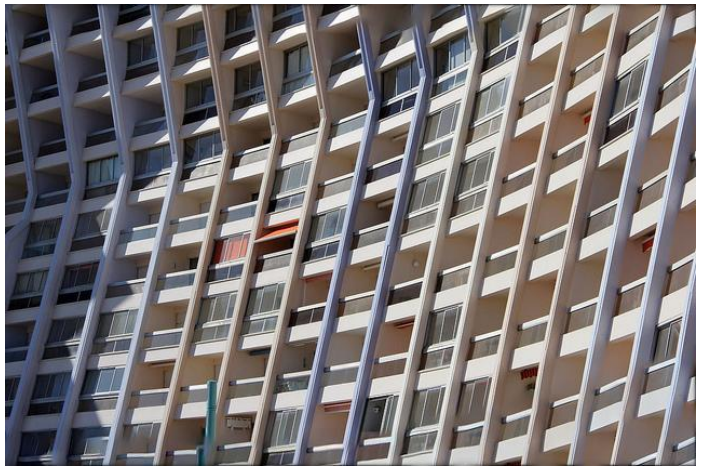
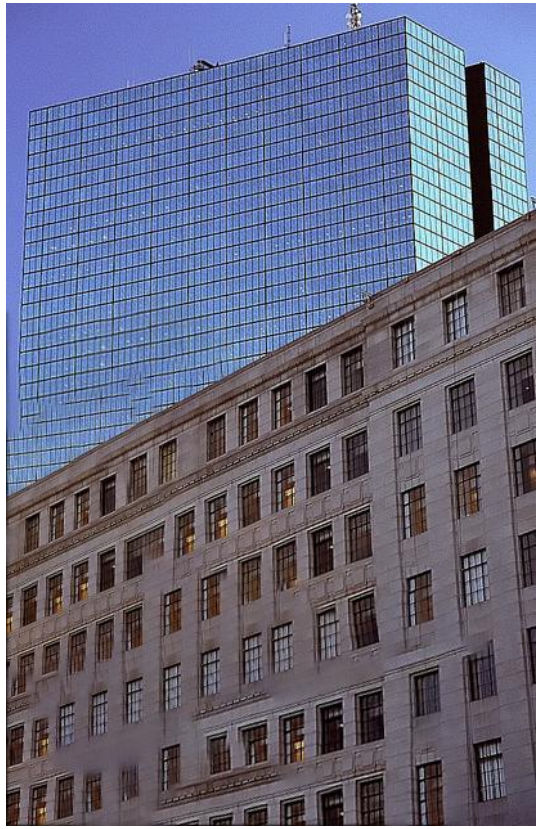
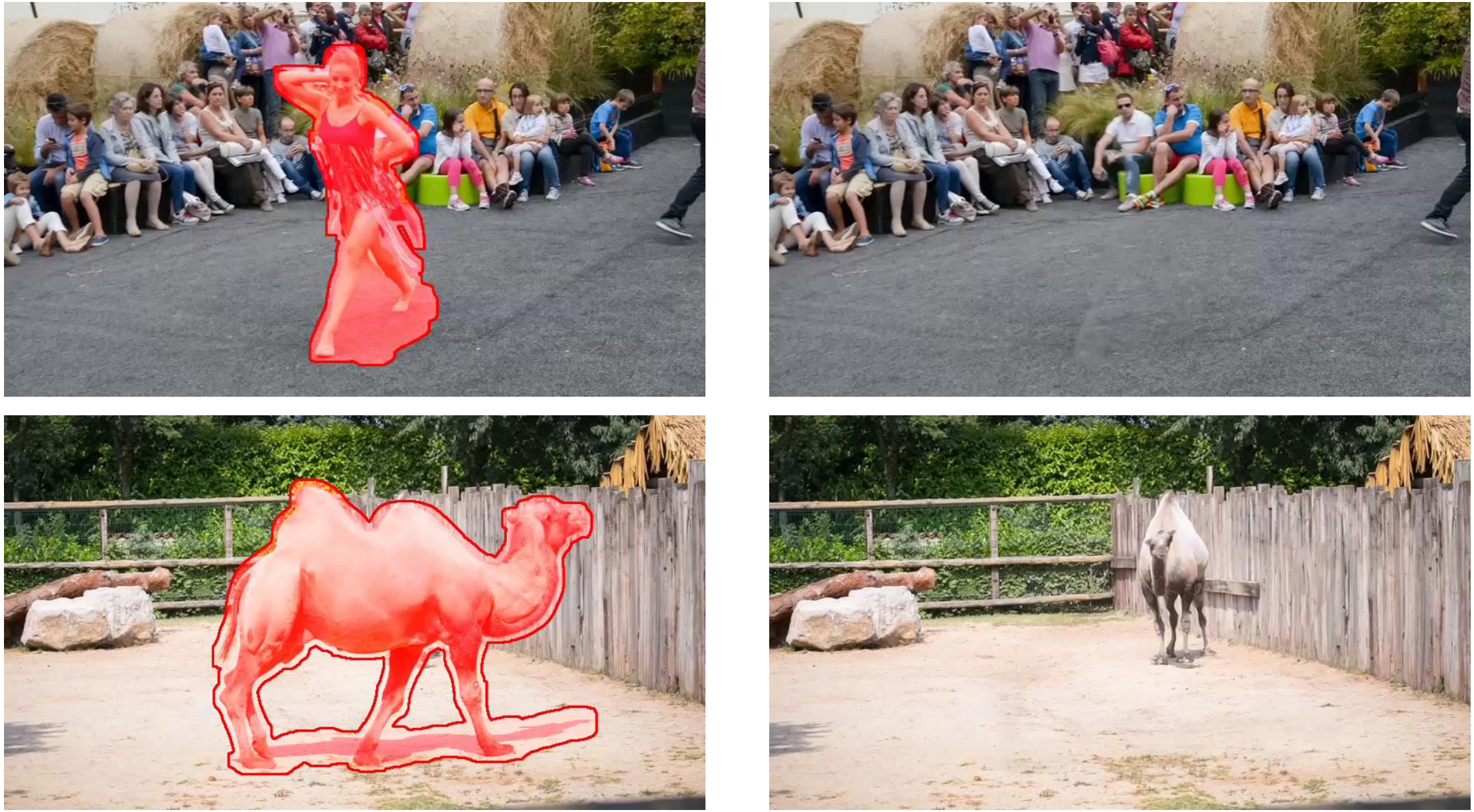





Image Completion [SIGGRAPH14]

- Revealing unseen pixels



Video Completion [SIGGRAPH Asia16]

- Revealing temporally coherent pixels

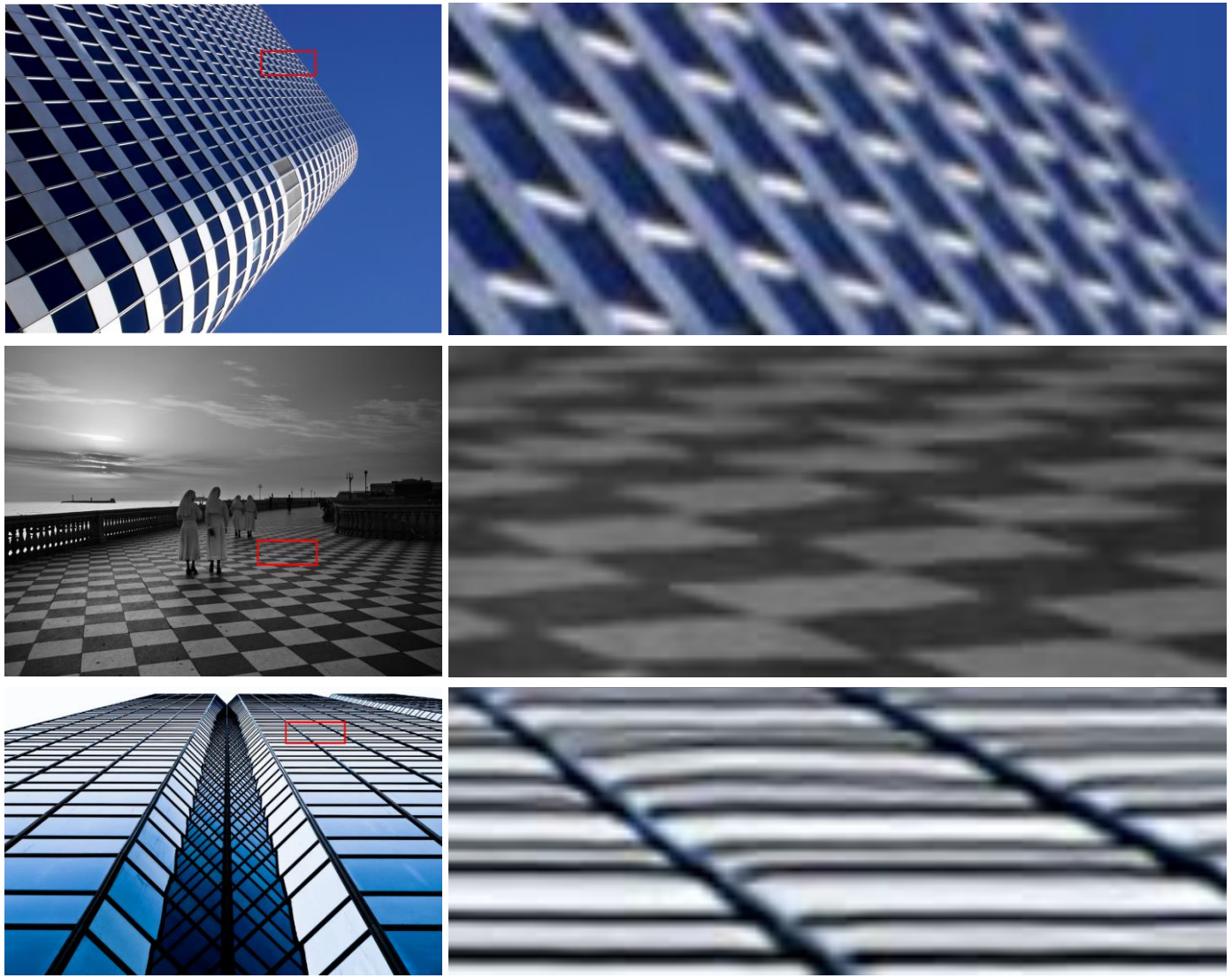


Image super-resolution [CVPR15]

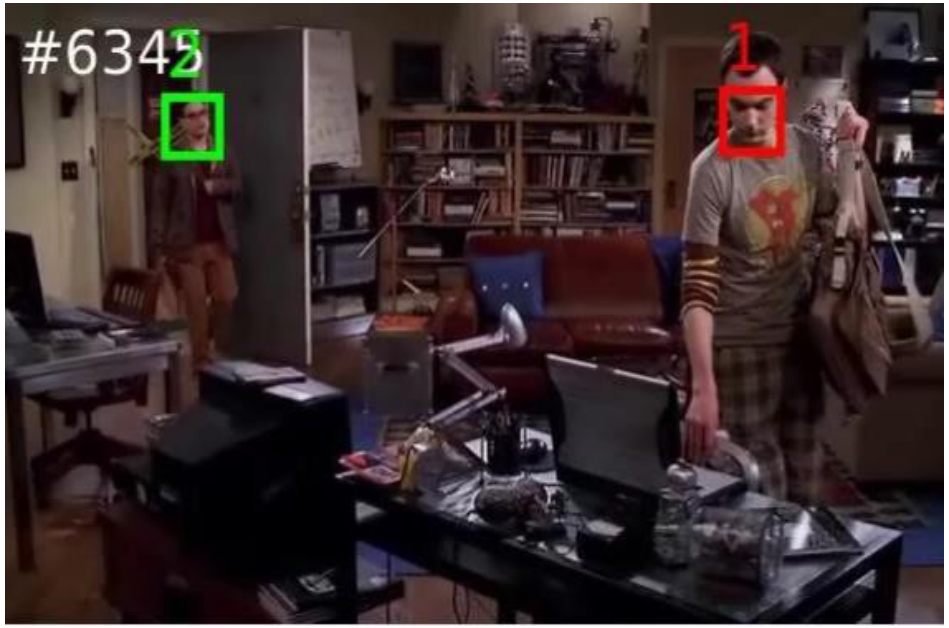
- Revealing unseen high frequency details

Depth upsampling | Noise reduction | Inverse halftoning | Texture removal

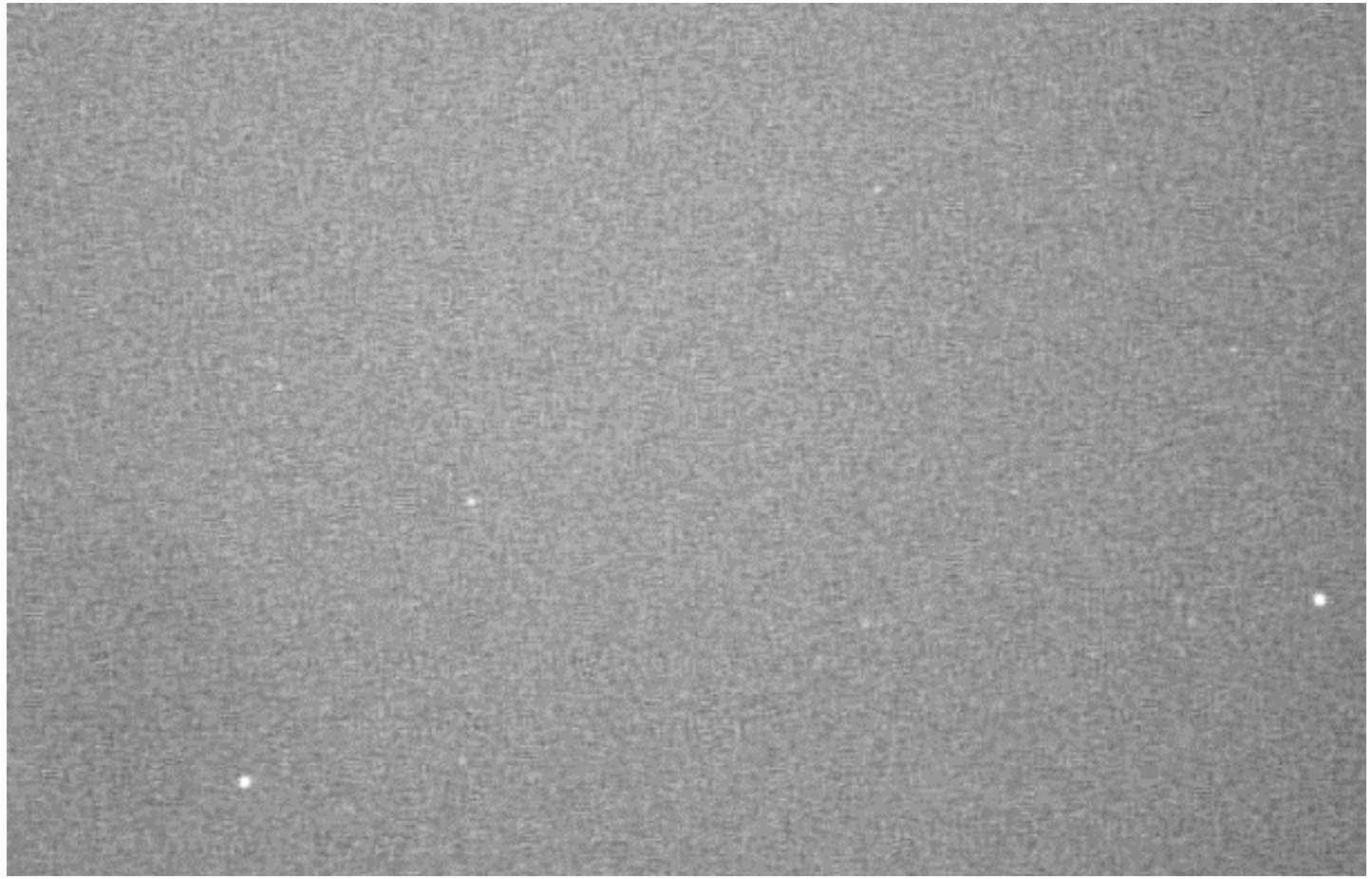
Deep Joint Image Filtering [ECCV16]

- Transferring structural details

Object tracking [ICCV15]



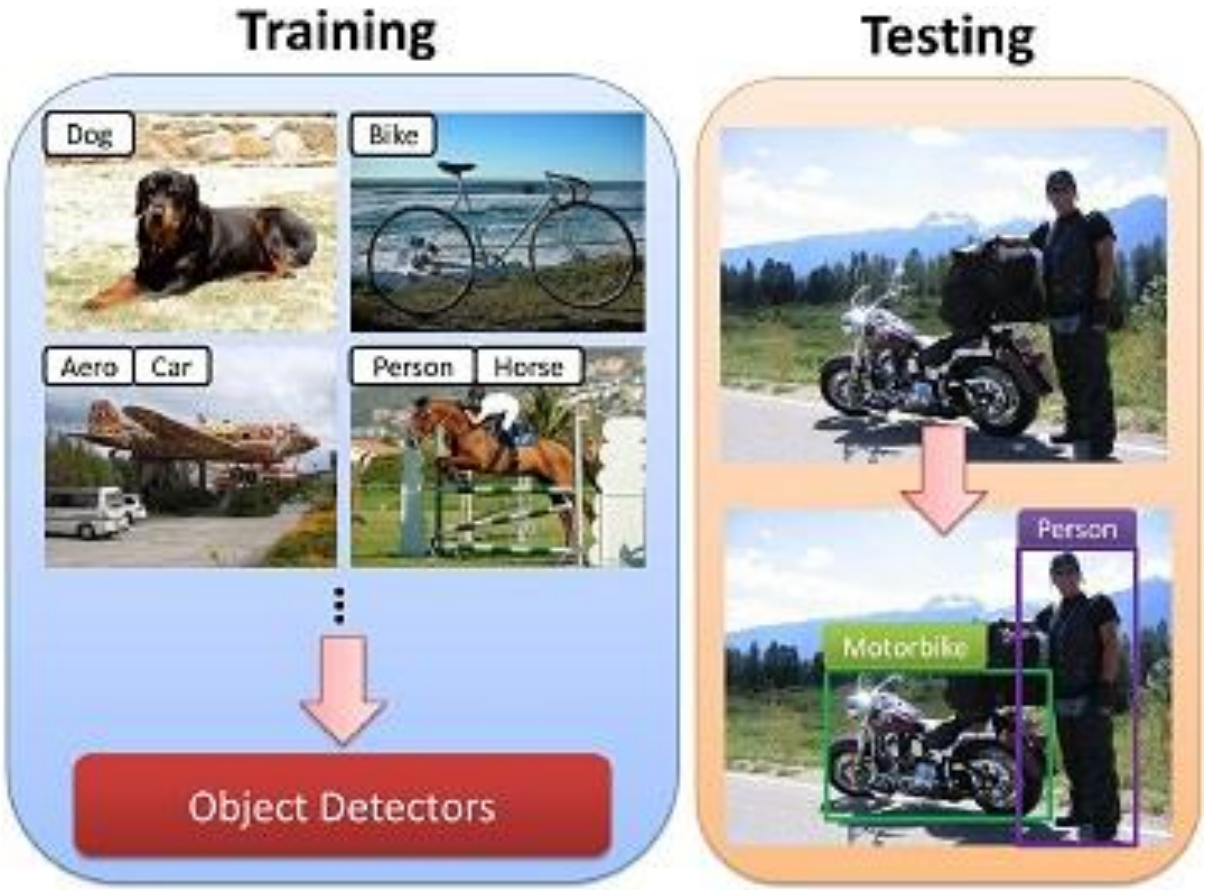
Multi-face tracking [ECCV16]



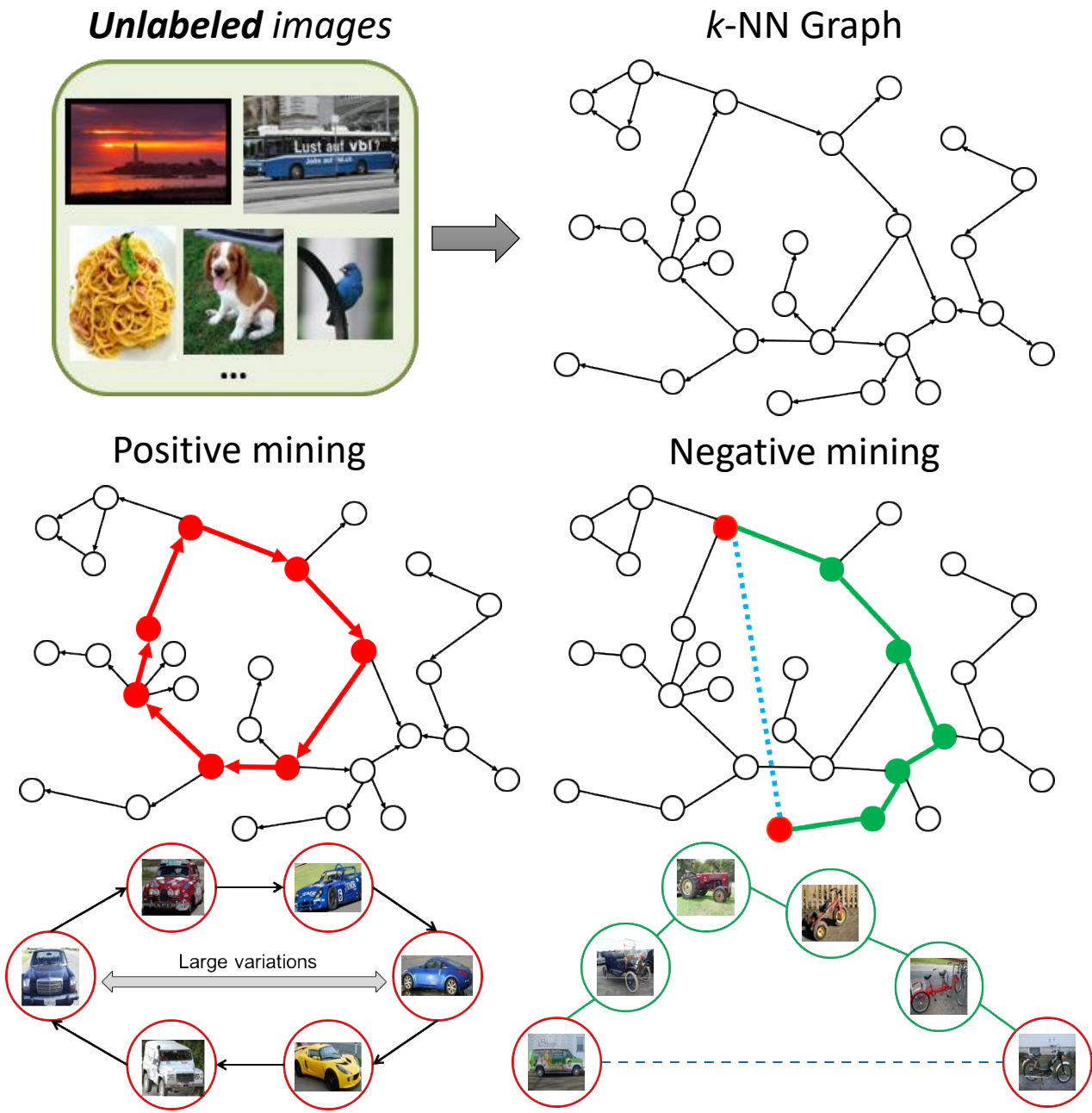
Detecting migrating birds [CVPR16]

Visual Tracking

- Locating moving objects across video frames



Weakly supervised localization [CVPR16]



Unsupervised feature learning [ECCV16]

Learning with weak labels

Teaching Assistant: Yuliang Zou

- 1st year PhD student in ECE, VT

- Email: ylzou@vt.edu
- Web: https://yuliang-zou.github.io/
- Office hour:
  - TBD
- Research:

A little about you

- Find a partner.
- Introduce yourself
  - Name?
  - Department?
  - Why are you taking this course?
  - One interesting fact?
- 3 mins

- **Introduce your partner!**

What is Computer Vision?

- Make computers understand images and videos.

- What kind of scene?
- Where are the cars?
- How far is the building?

What is Computer Vision?

- Make computers understand images and videos.

- What are they doing?
- Why is this happening?
- What is important?
- What will I see?

Computer Vision and Nearby Fields

Digital Image Processing
Computational Photography

Images (2D)

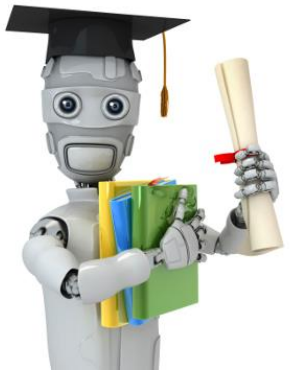
Computer Vision

Computer Graphics

Geometry (3D)
Shape

+

Photometry
Appearance



Machine learning:
Vision = Machine learning applied to visual data

Visual data on the Internet

- Flickr
  - 10+ billion photographs
  - 60 million images uploaded a month
- Facebook
  - 250 billion+
  - 300 million a day
- Instagram
  - 55 million a day
- YouTube
  - 100 hours uploaded every minute

CISCO **90%** of net traffic will be visual!

Mostly about cats

Too big for humans

http://www.petittube.com/

- Need automatic tools to access and analyze visual data!

Vision is Really Hard

- Vision is an amazing feature of natural intelligence
  - Visual cortex occupies about 50% of Macaque brain
  - More human brain devoted to vision than anything else

Why is Computer Vision Hard?

Why is Computer Vision Hard?

What did you see?

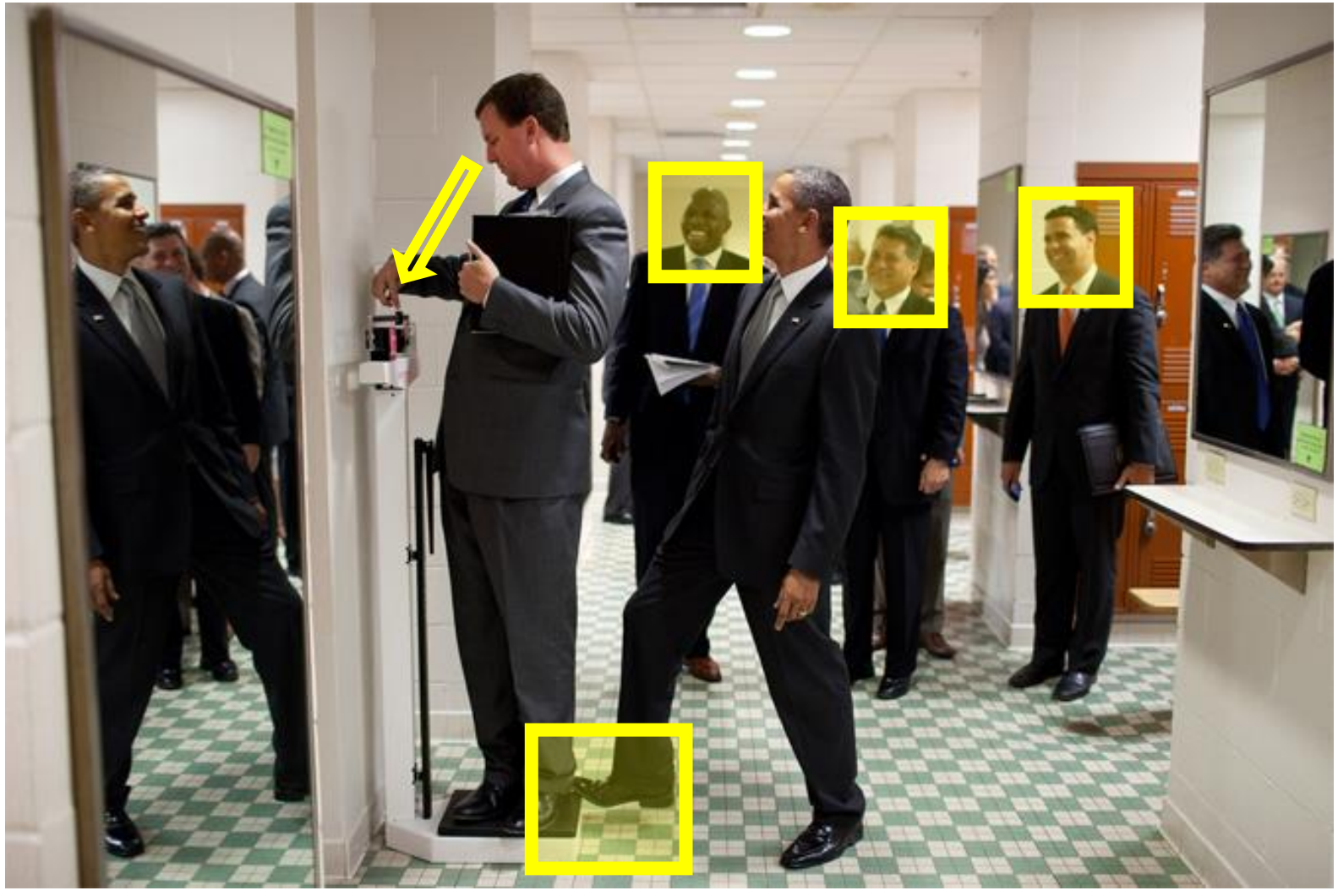
- Where this picture was taken?
- How many people are there?
- What are they doing?
- What object the person on the left standing on?
- Why this is a funny picture?

Why is Computer Vision Hard?

Why is Computer Vision Hard?

Why is Computer Vision Hard?

Why is Computer Vision Hard?



Why is Computer Vision Hard?

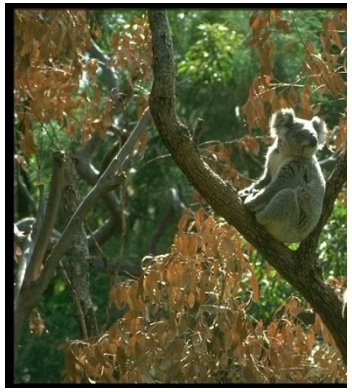
Why is Computer Vision Hard?

Computer: okay, it’s a funny picture

Challenges: Many nuisance parameters

Illumination

Object pose



Clutter

Occlusions

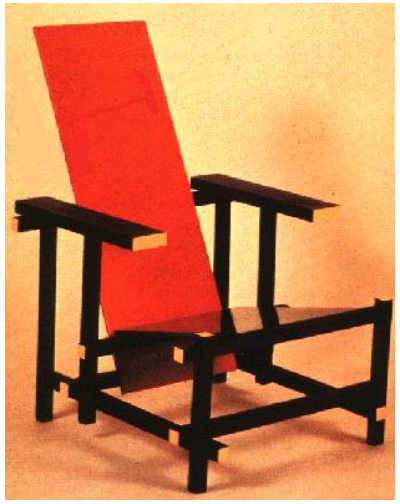
Intra-class appearance

Viewpoint

Slide credit: Kristen Grauman

Challenges: Intra-class variation



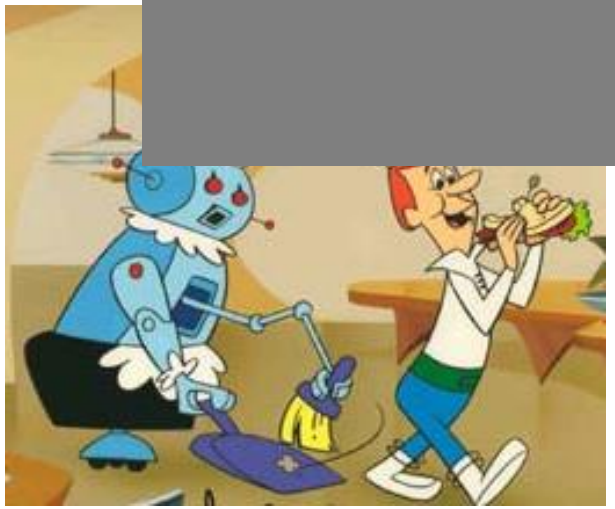



Slide credit: Fei-Fei, Fergus & Torralba

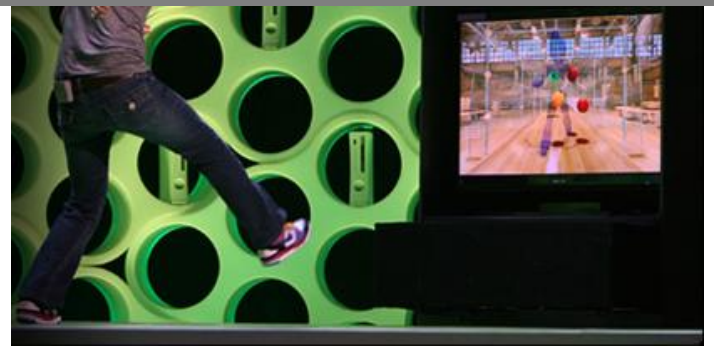
Challenges: Importance of context

Slide credit: Fei-Fei, Fergus & Torralba

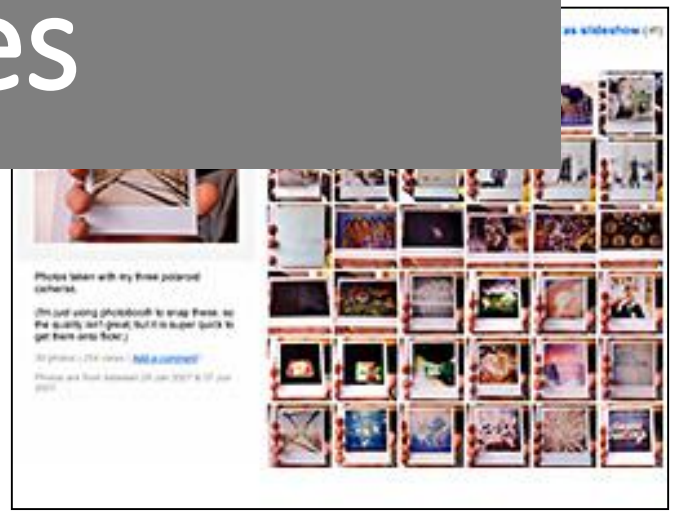
Computer Vision Technology Can Better Our Lives



Comfort



Fun



Access

History of Computer Vision

Marvin Minsky, MIT
Turing award, 1969

"In 1966, Minsky hired a first-year undergraduate student and assigned him a problem to solve over the summer:

connect a camera to a computer and get the machine to describe what it sees."

Crevier 1993, pg. 88

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence Group
Vision Memo. No. 100.

July 7, 1966

Half a century later,
we're still working on it.

THE SUMMER VISION PROJECT

Seymour Papert

The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".

History of Computer Vision

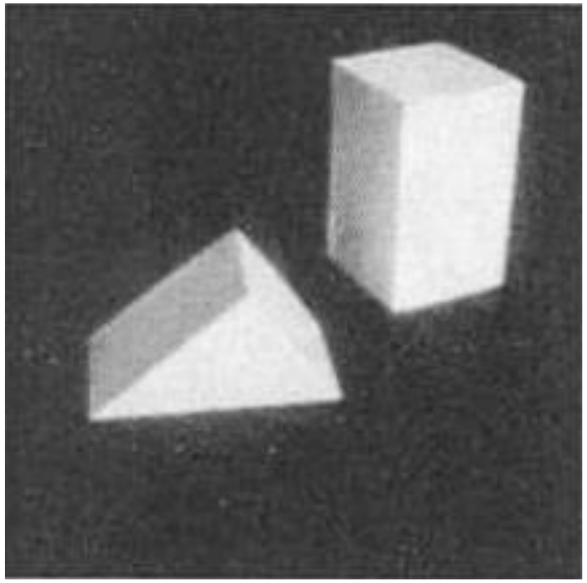
Marvin Minsky, MIT
Turing award, 1969

Gerald Sussman, MIT

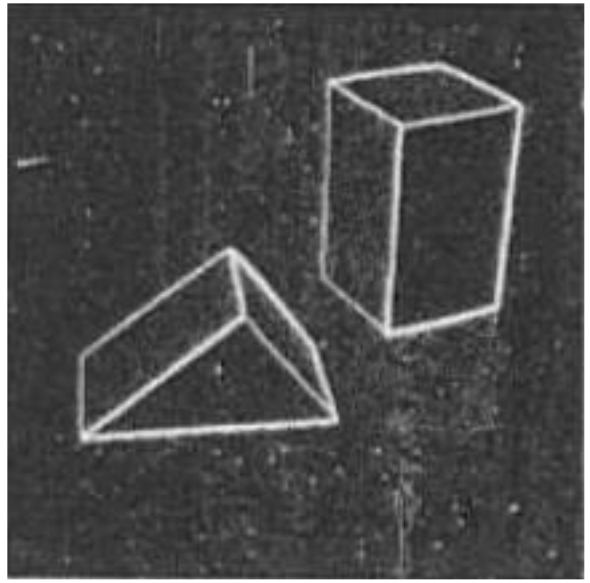
"You'll notice that Sussman never worked in vision again!" – Berthold Horn

1960's: interpretation of synthetic worlds

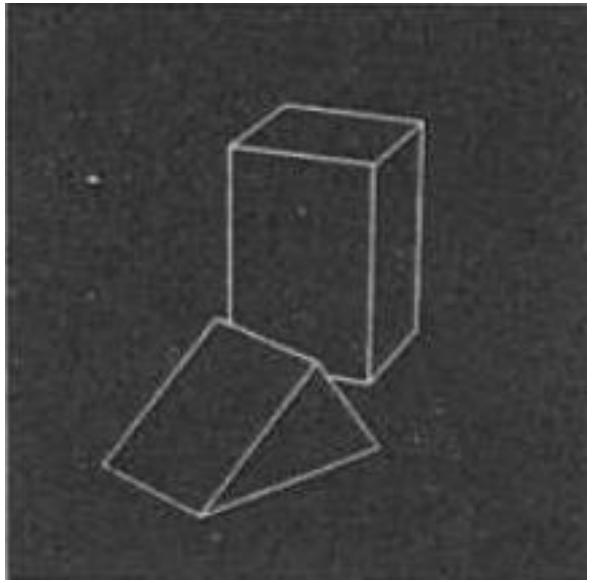
Larry Roberts
"Father of Computer Vision"



Input image



2x2 gradient operator

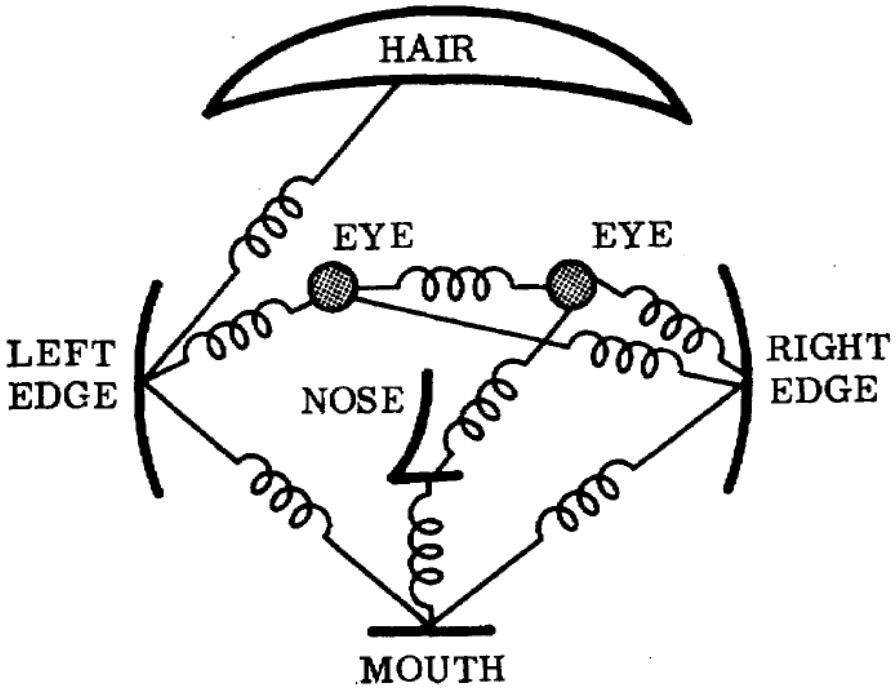


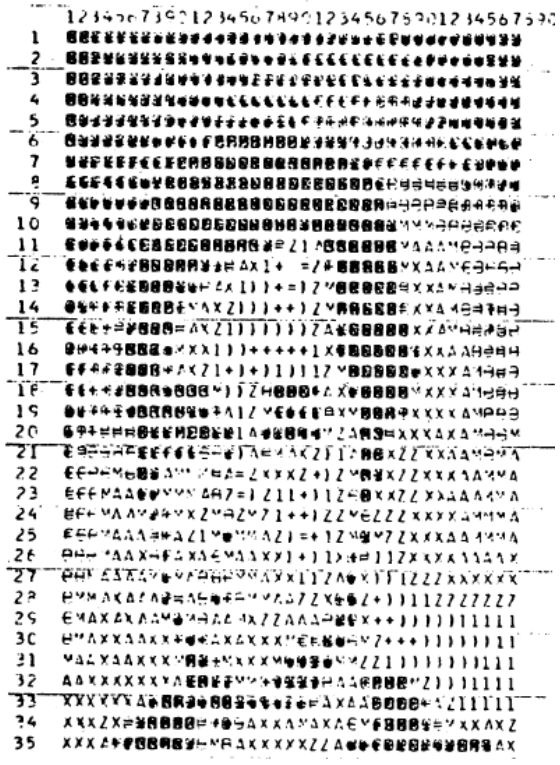
computed 3D model
rendered from new viewpoint

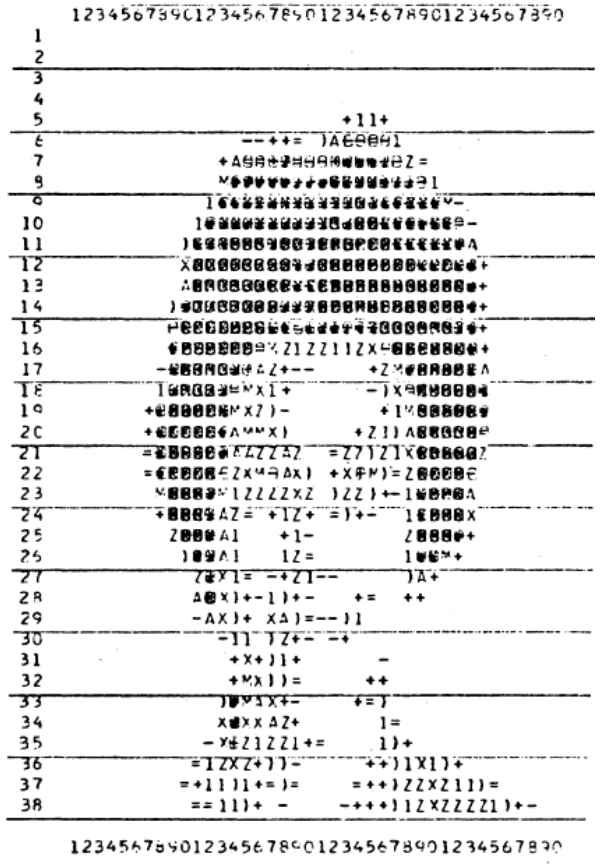
Larry Roberts PhD Thesis, MIT, 1963,
Machine Perception of Three-Dimensional Solids

Slide credit: Steve Seitz

1970’s: some progress on interpreting selected images

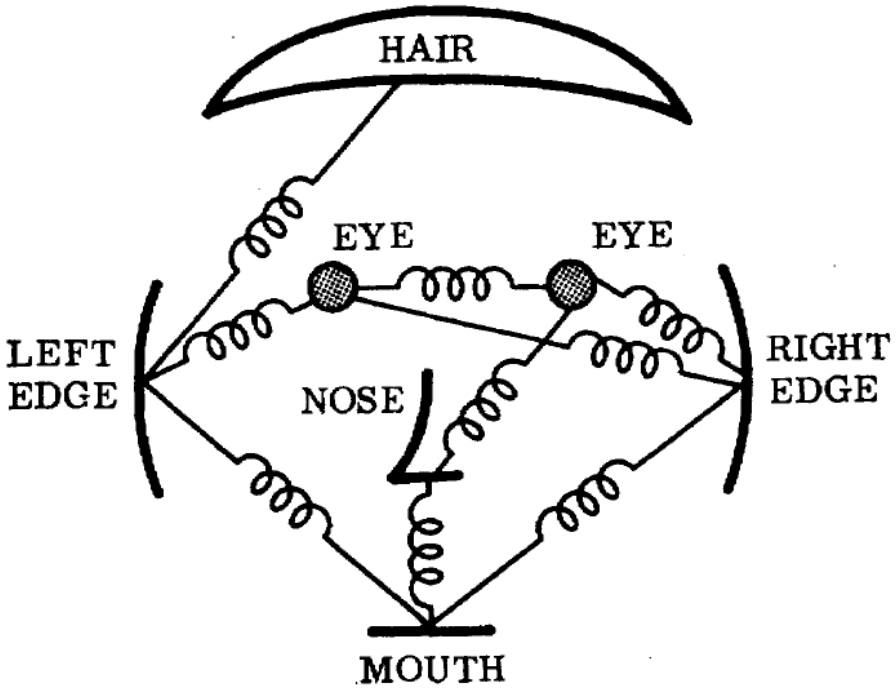


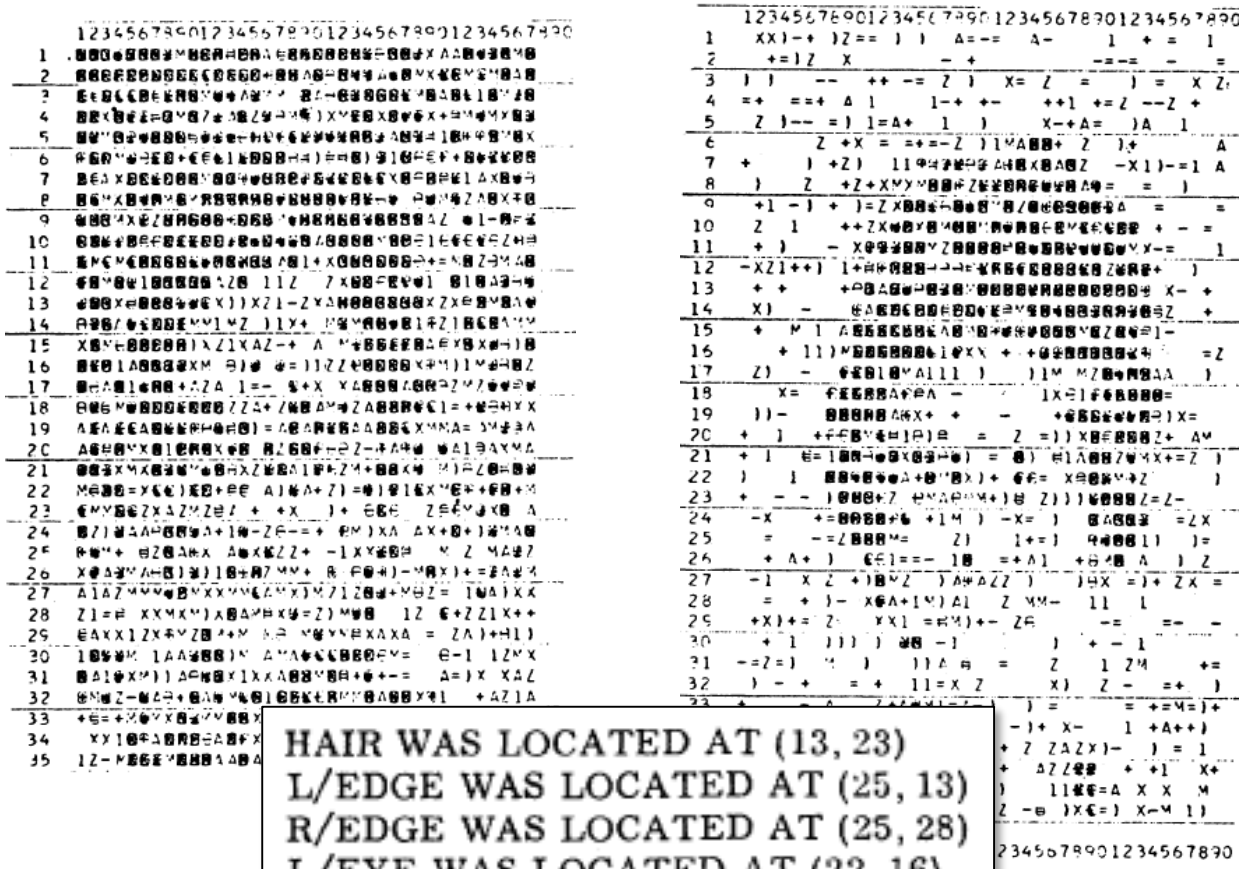




The representation and matching of pictorial structures
Fischler and Elschlager, 1973

1970’s: some progress on interpreting selected images



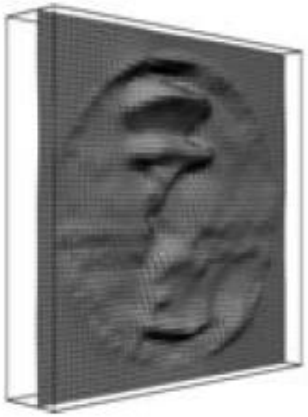


HAIR WAS LOCATED AT (13, 23)
L/EDGE WAS LOCATED AT (25, 13)
R/EDGE WAS LOCATED AT (25, 28)
L/EYE WAS LOCATED AT (22, 16)
R/EYE WAS LOCATED AT (22, 23)
NOSE WAS LOCATED AT (27, 20)
MOUTH WAS LOCATED AT (29, 19)

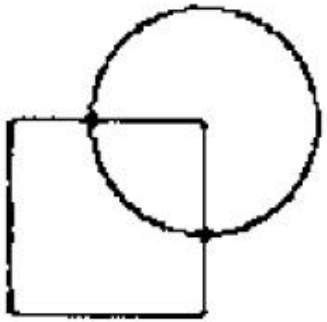
The representation and matching of pictorial structures
Fischler and Elschlager, 1973

1980's: ANNs come and go; shift toward geometry and increased mathematical rigor

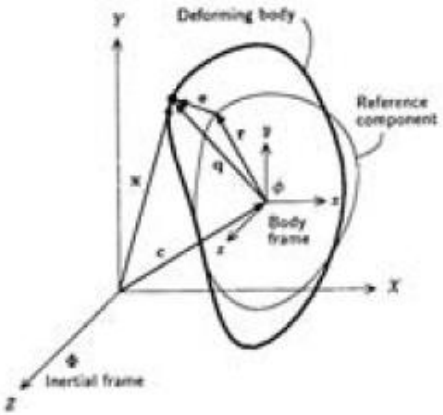
(a)



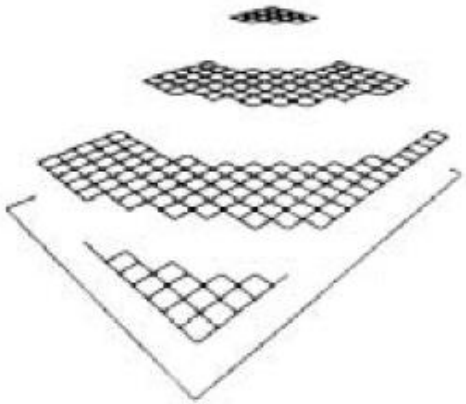
(b)



(c)



(d)



(e)

(f)

Image credit: Rick Szeliski

Goodbye science

1990’s: face recognition; statistical analysis in vogue

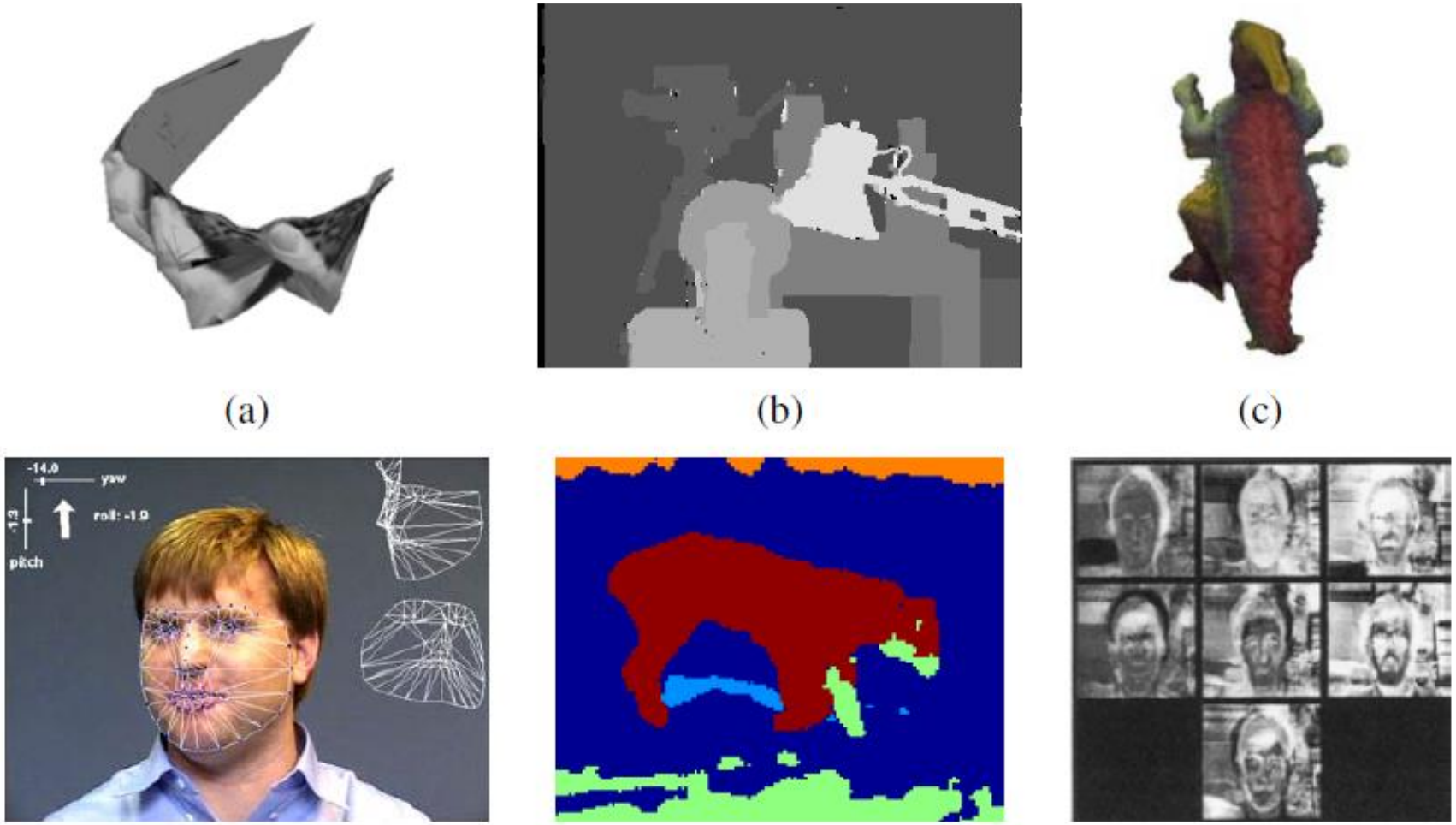
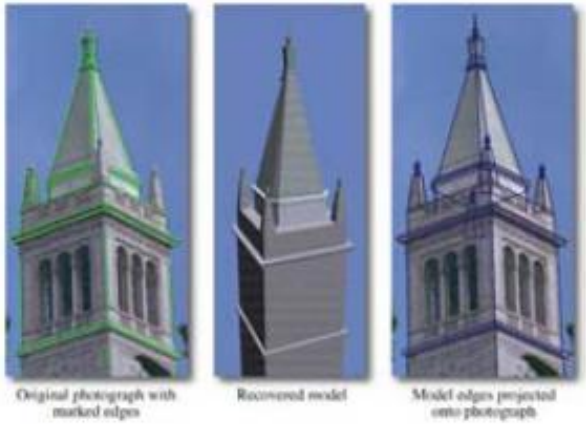


Image credit: Rick Szeliski

2000's: broader recognition; large annotated datasets available; video processing starts

(a)



(b)

(c)

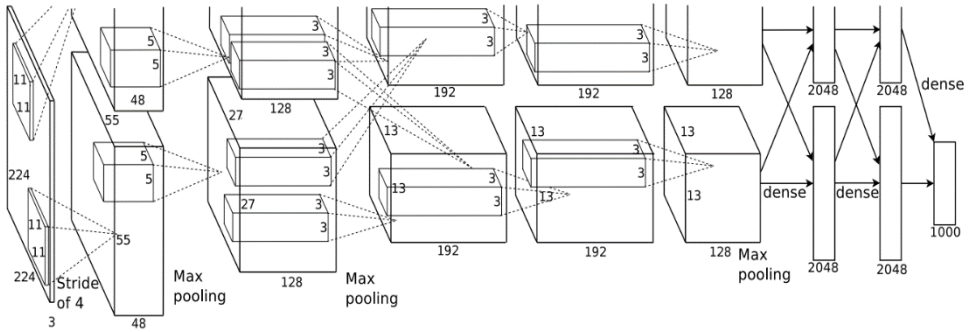
(d)

(e)

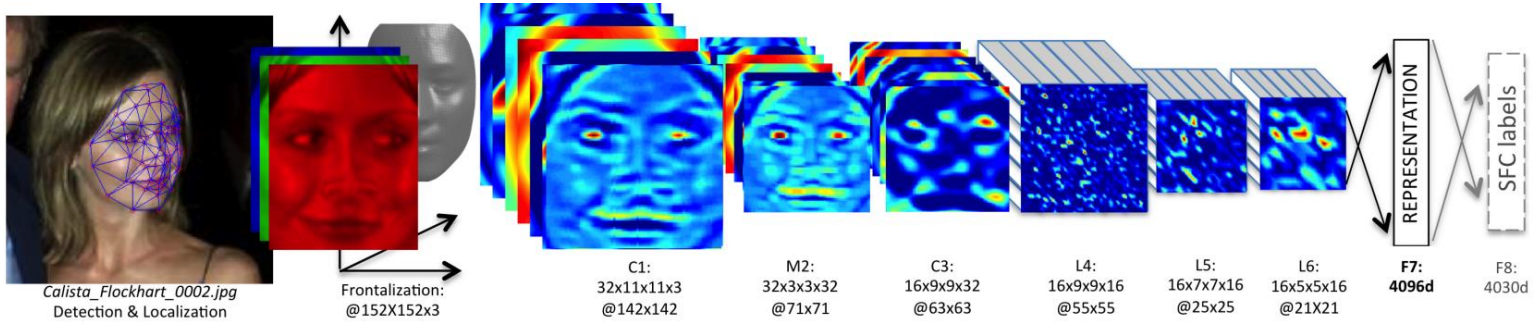
(f)

Image credit: Rick Szeliski

2010's: resurgence of deep learning

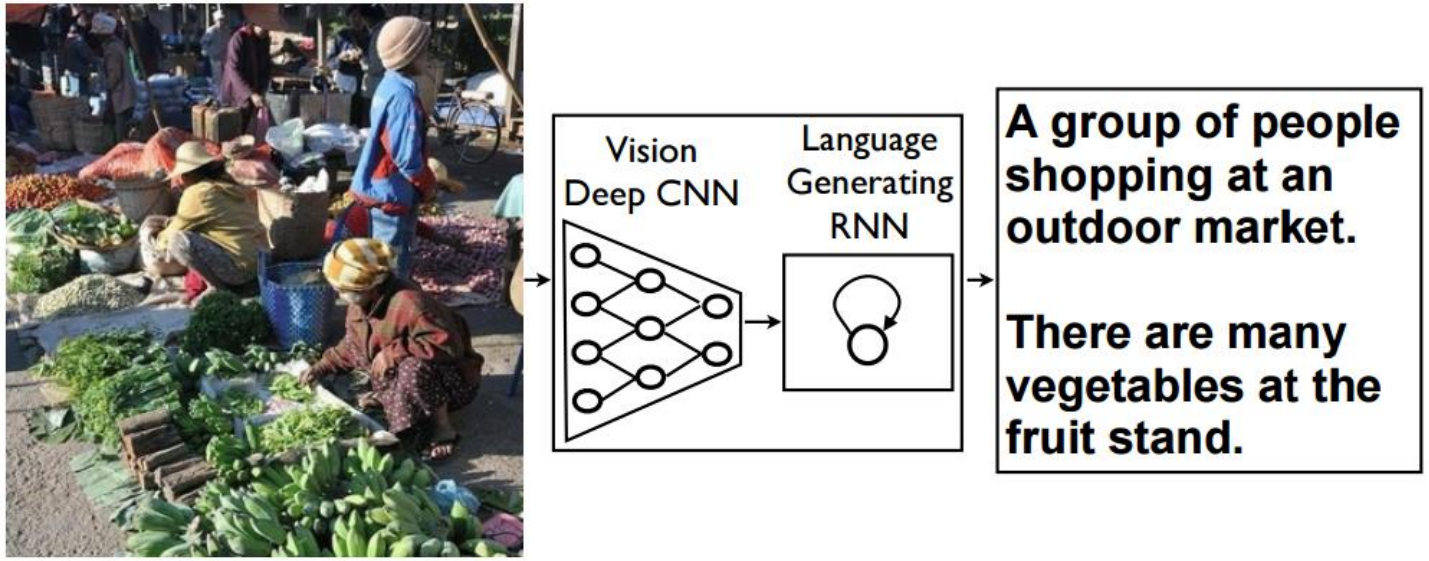


[AlexNet NIPS 2012]



[DeepFace CVPR 2014]

[DeepPose CVPR 2014]



[Show, Attend and Tell ICML 2015]

2020’s: autonomous vehicles

2030’s: robot uprising?

Examples of Computer Vision Applications

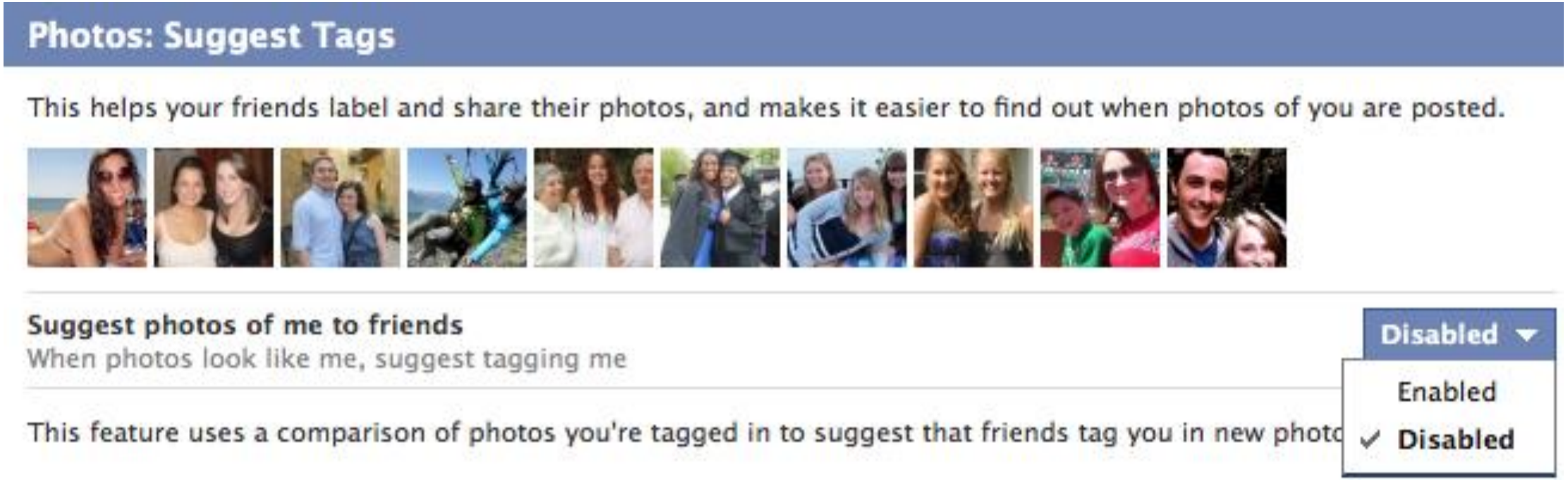
- How is computer vision used today?

Face detection

- Most digital cameras and smart phones detect faces (and more)
  - Canon, Sony, Fuji, …
- For smart focus, exposure compensation, and cropping

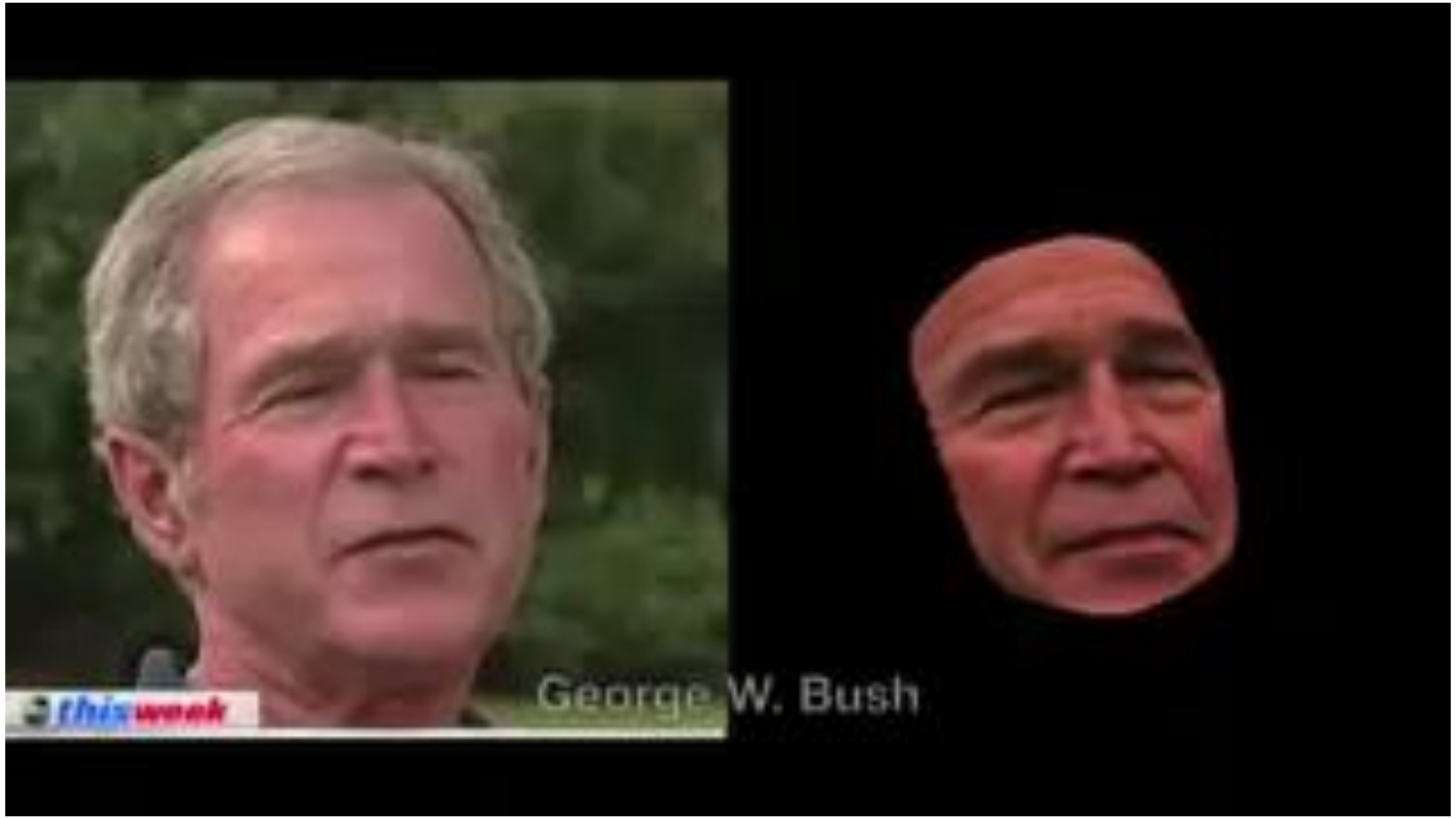
Slide credit: Steve Seitz

Face recognition



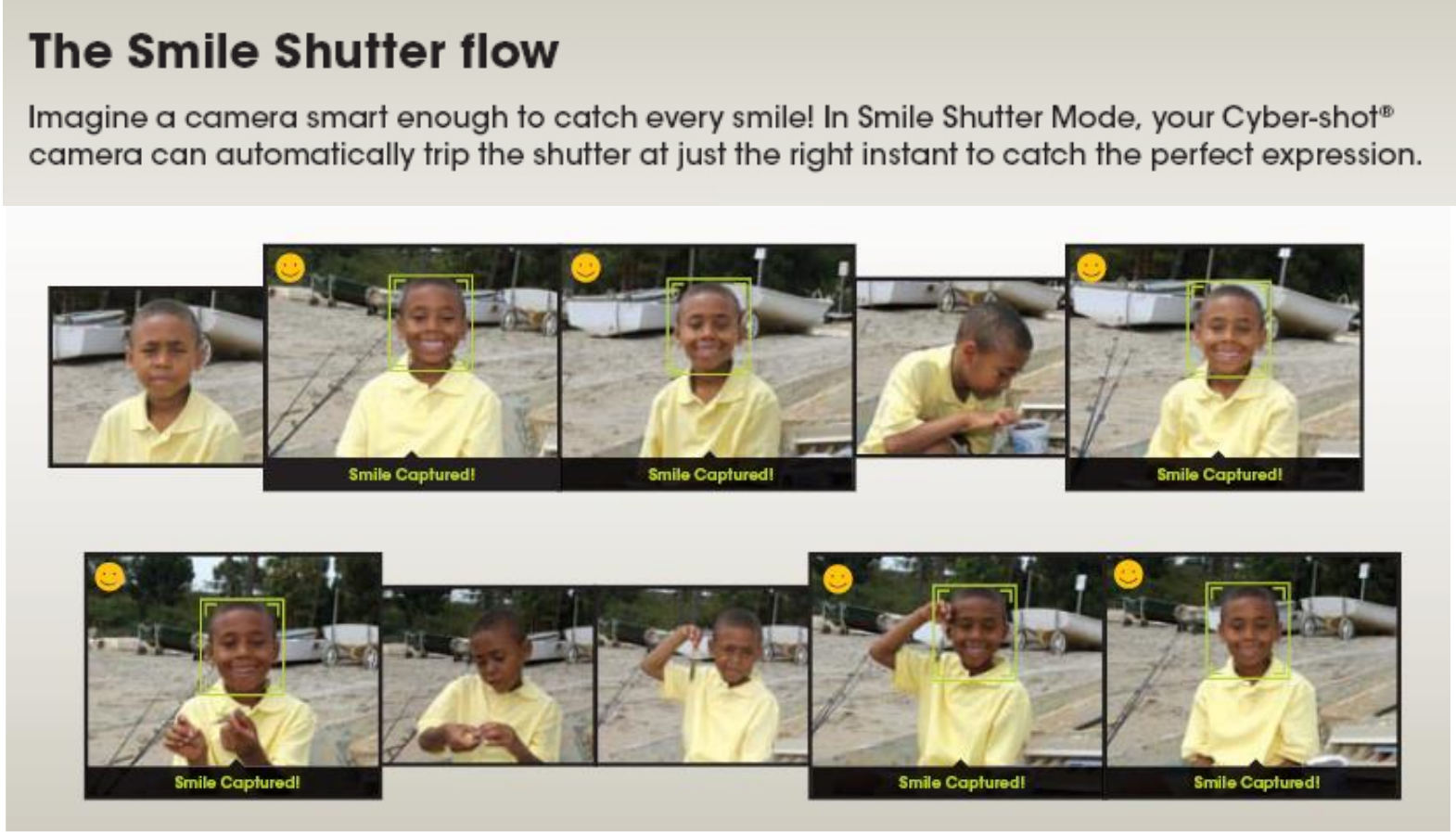
Facebook face auto-tagging

Face Landmark Alignment – 3D Persona



What Makes Tom Hanks Look Like Tom Hanks ICCV 2015

Smile Detection

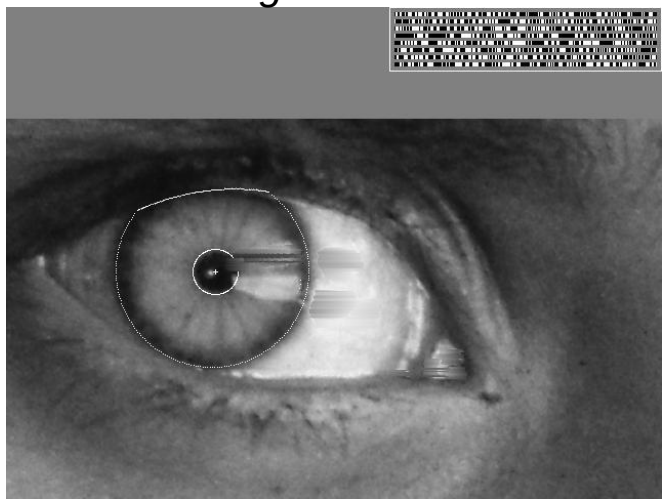


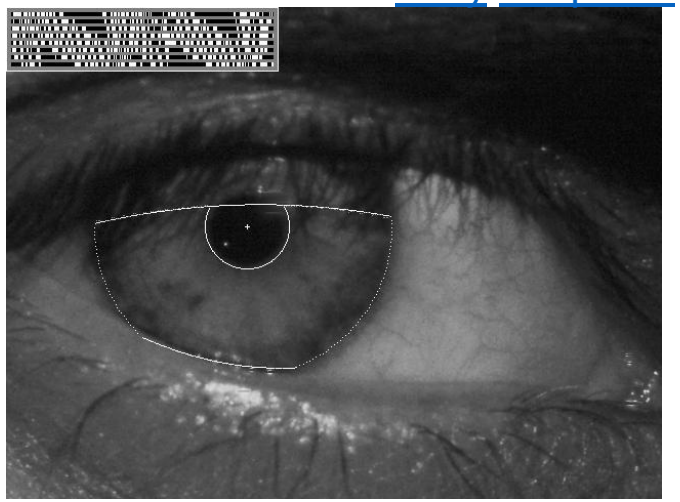
Sony Cyber-shot® T70 Digital Still Camera

Slide credit: Steve Seitz

Vision-based Biometrics

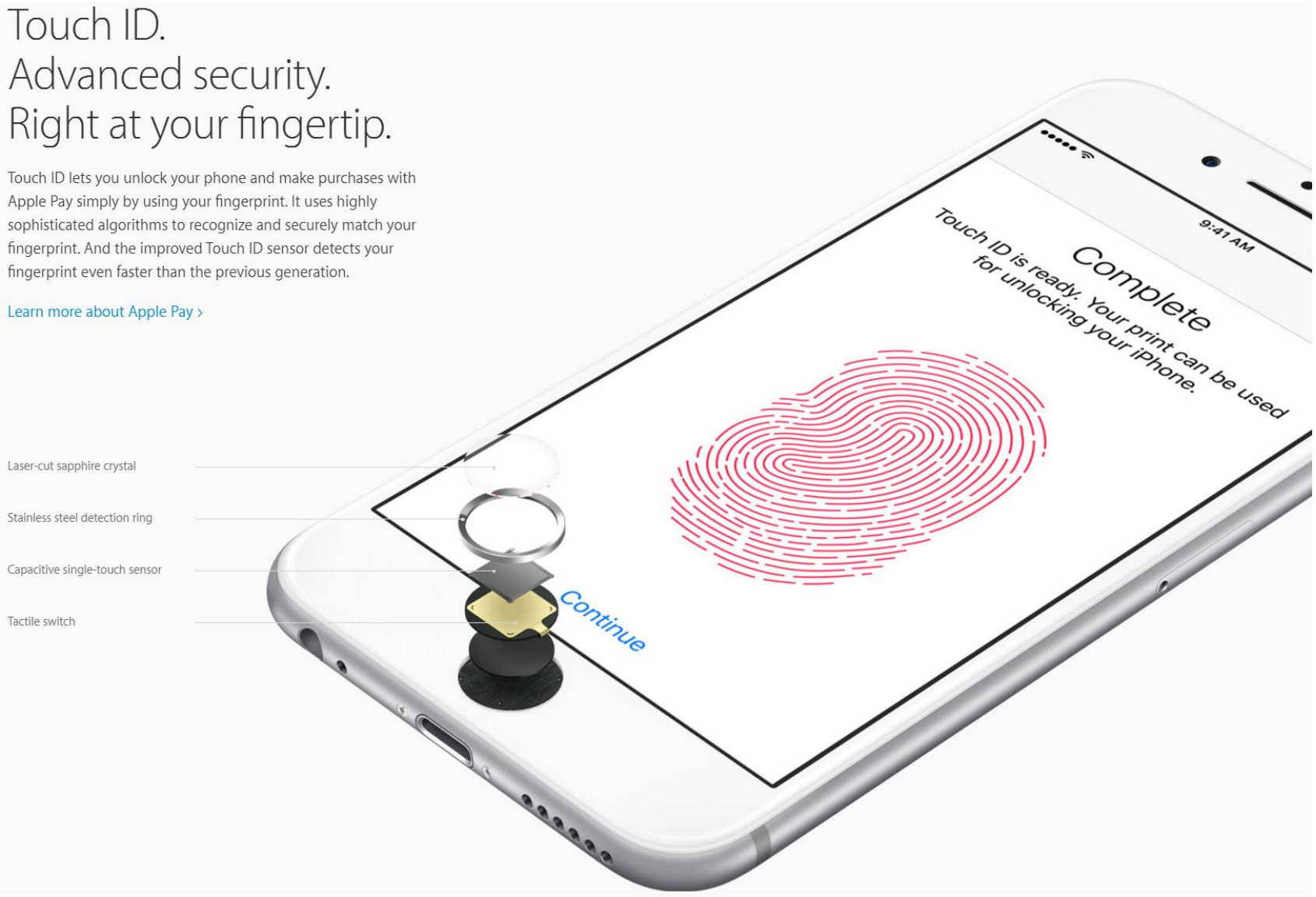
"*How the Afghan Girl was Identified by Her Iris Patterns*" Read the story wikipedia





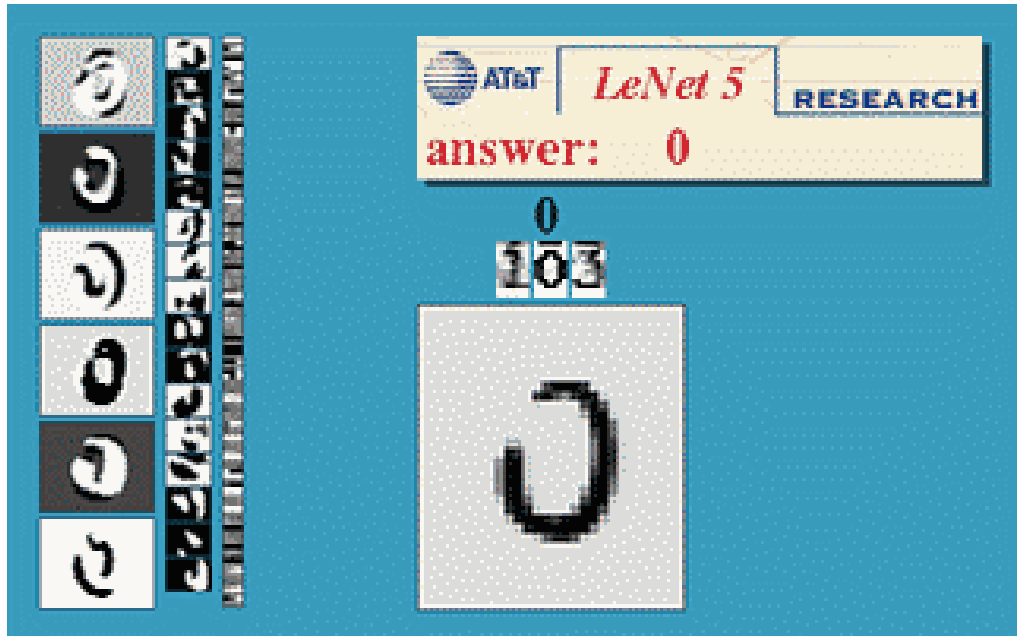
Slide credit: Steve Seitz

Vision-based Biometrics



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
http://www.research.att.com/~yann/

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

Slide credit: Steve Seitz

Computer vision in sports

Hawk-Eye: helping/improving referee decisions

Computer vision in sports

SportVision: improving viewer experiences

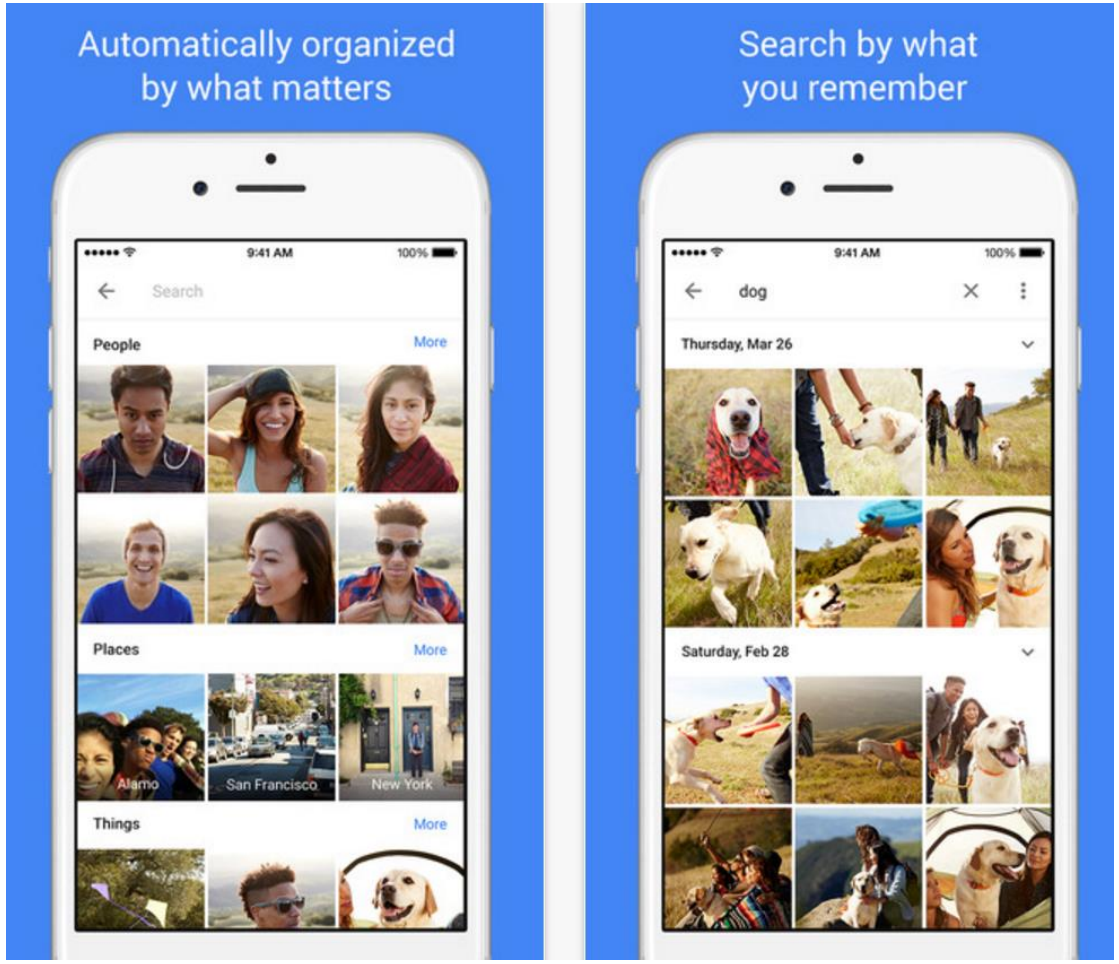
Computer vision in sports

Replay Technologies: improving viewer experiences

Computer vision in sports

Play tracking

Visual recognition for photo organization



Google photo

Earth viewers (3D modeling)

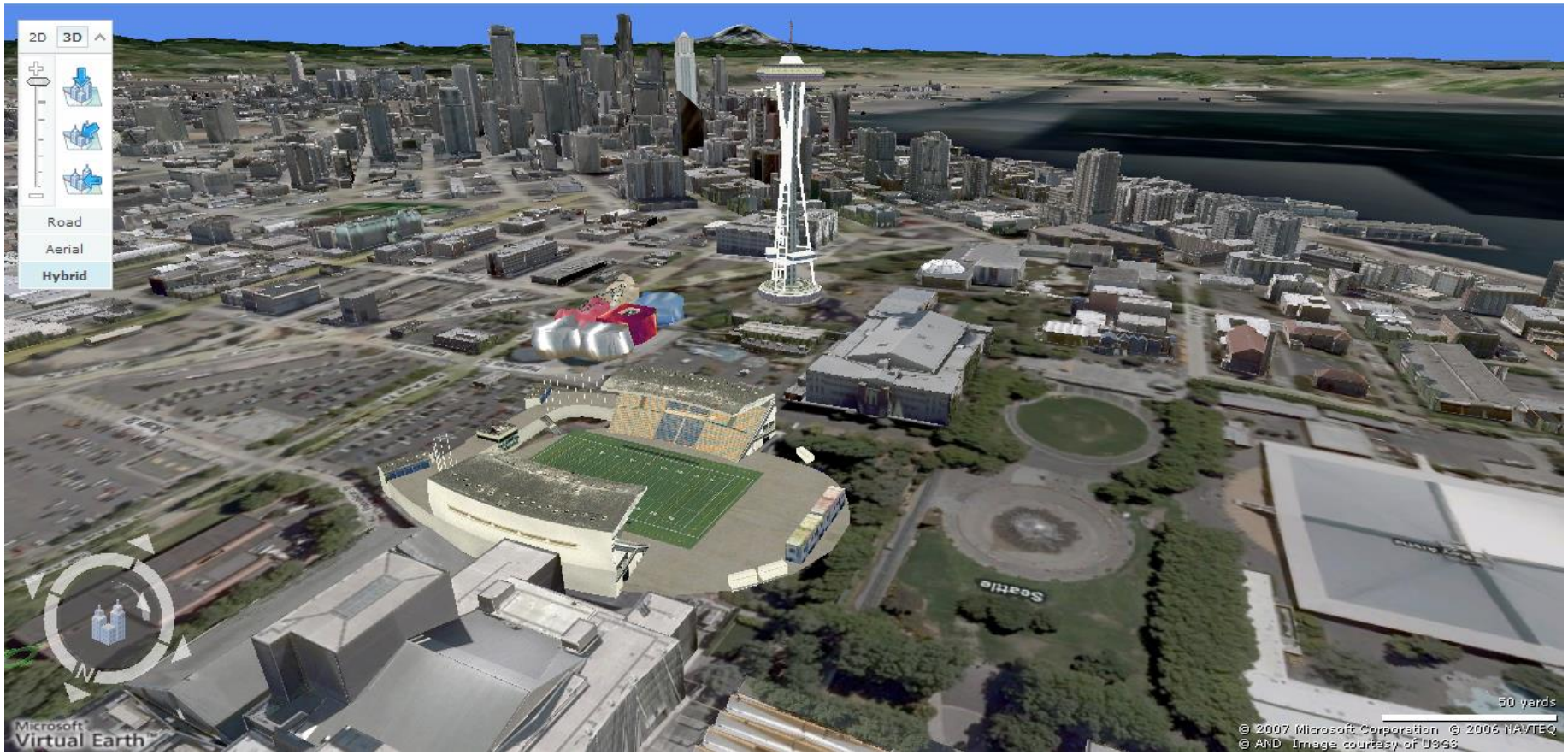


Image from Microsoft's Virtual Earth
(see also: Google Earth)

Slide credit: Steve Seitz

3D from thousands of images

[Furukawa et al. CVPR 2010]

Microsoft PhotoSynth: Photo Tourism

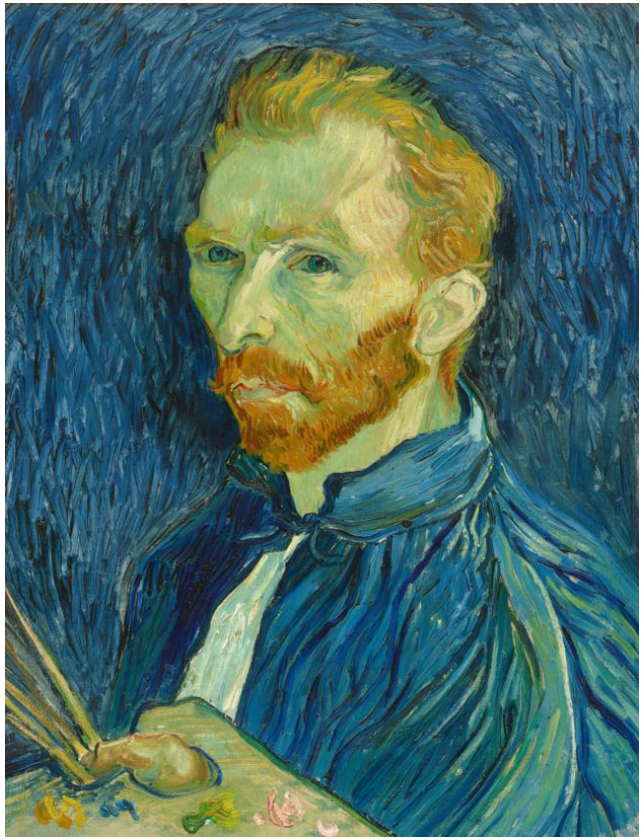
First-person Hyperlapse Videos

[Kopf et al. SIGGRAPH 2014]

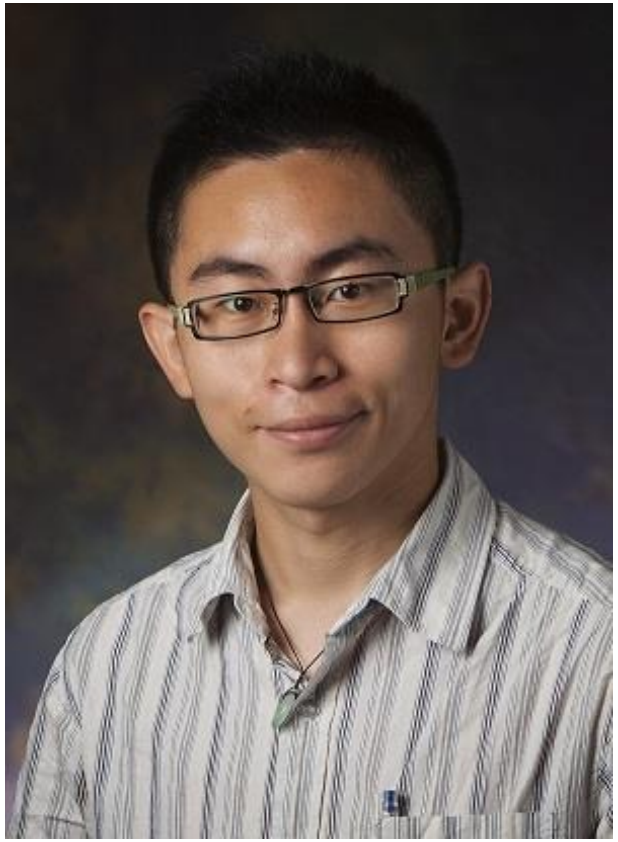
3D Time-lapse from Internet Photos

3D Time-lapse from Internet Photos, ICCV 2015

Style transfer



Source image (**Style**)



Target image (**Content**)

Output (deepart)

A Neural Algorithm of Artistic Style [Gatys et al. 2015]

Special effects: Matting and composition

Kylie Minogue - Come Into My World

Special effects: Shape capture

The Matrix movies, ESC Entertainment, XYZRGB, NRC

Slide credit: Steve Seitz

Special effects: Motion capture

Pirates of the Carribean, Industrial Light and Magic

Slide credit: Steve Seitz

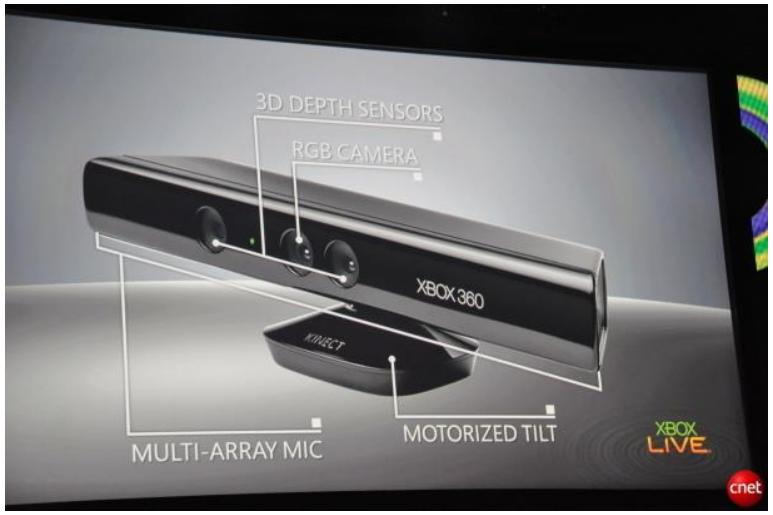
Google cars

Google in talks with Ford, Toyota and Volkswagen to realise driverless cars

http://www.theatlantic.com/technology/archive/2014/05/all-the-world-a-track-the-trick-that-makes-googles-self-driving-cars-work/370871/

Interactive Games: Kinect

- Object Recognition: http://www.youtube.com/watch?feature=iv&v=fQ59dXOo63o
- Mario: http://www.youtube.com/watch?v=8CTJL5lUjHg
- 3D: http://www.youtube.com/watch?v=7QrnwoO1-8A
- Robot: http://www.youtube.com/watch?v=w8BmgtMKFbY



Vision in space

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

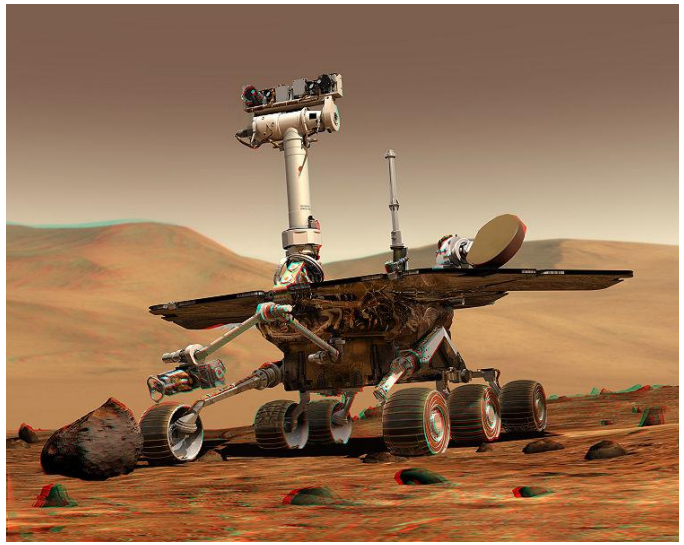
- Panorama stitching
- 3D terrain modeling
- Obstacle detection, position tracking
- For more, read "Computer Vision on Mars" by Matthies et al.

Industrial robots

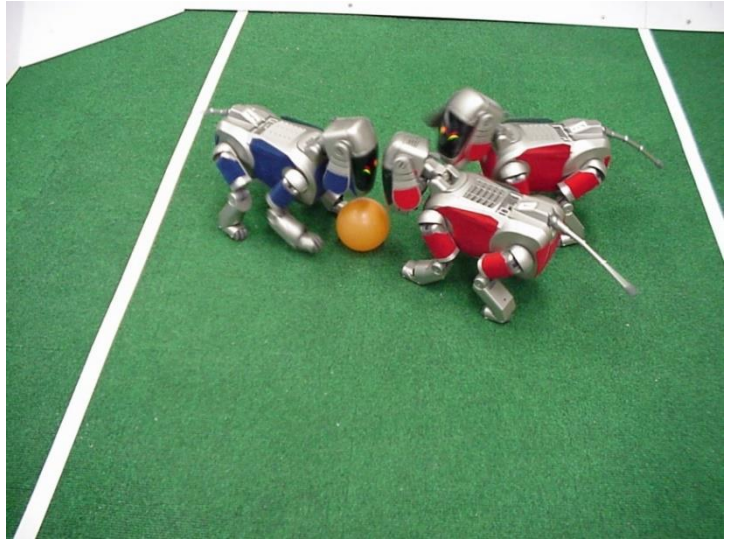
Vision-guided robots position nut runners on wheels

http://www.automationworld.com/computer-vision-opportunity-or-threat

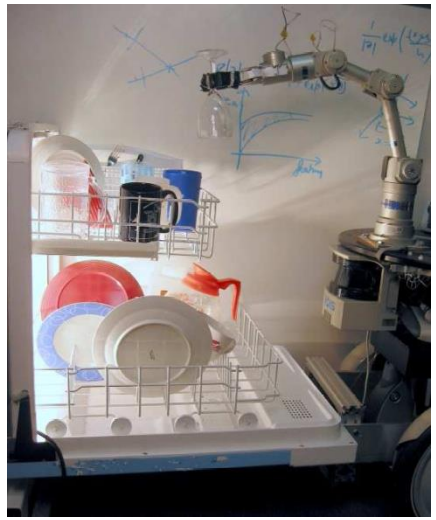
Mobile robots



NASA's Mars Spirit Rover



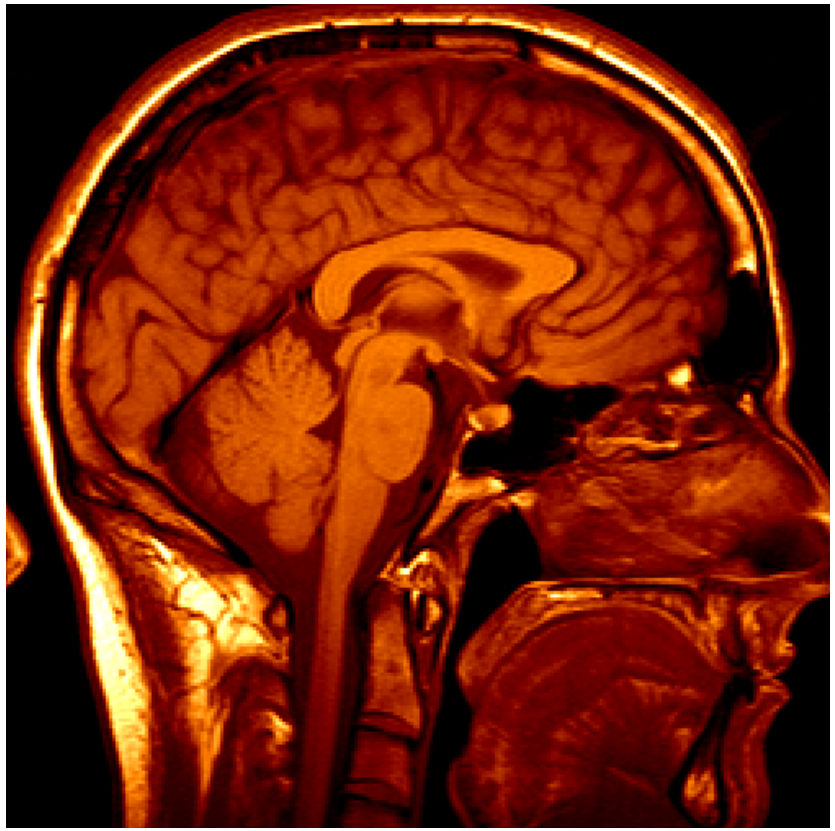
http://www.robocup.org/



Saxena et al. 2008
STAIR at Stanford

http://www.youtube.com/watch?v=DF39Ygp53mQ

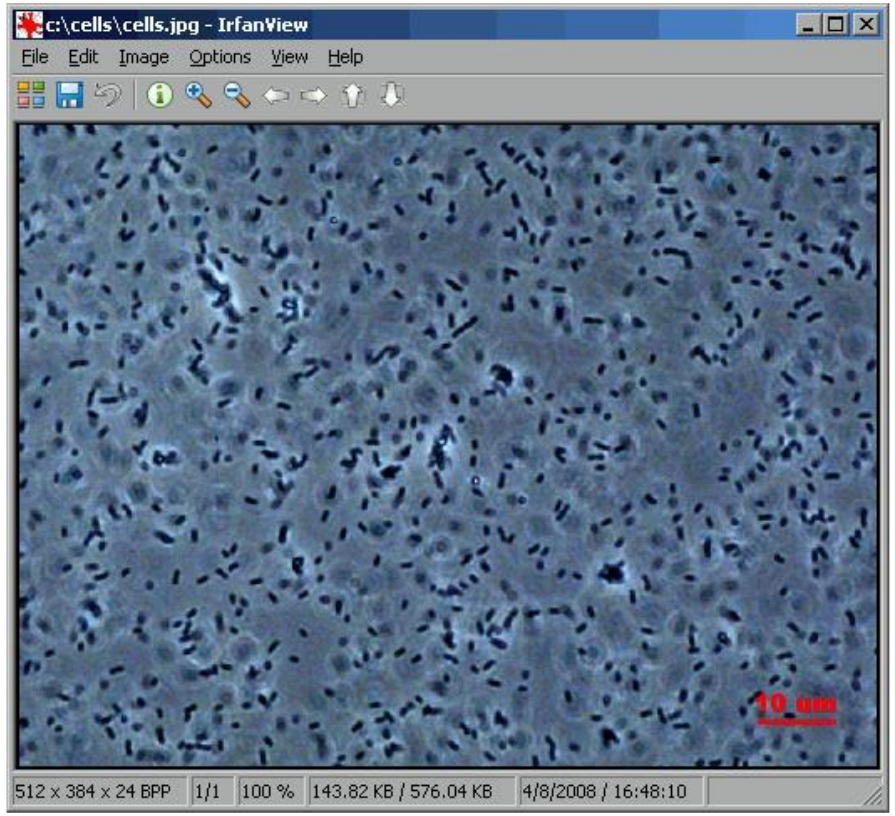
Medical imaging



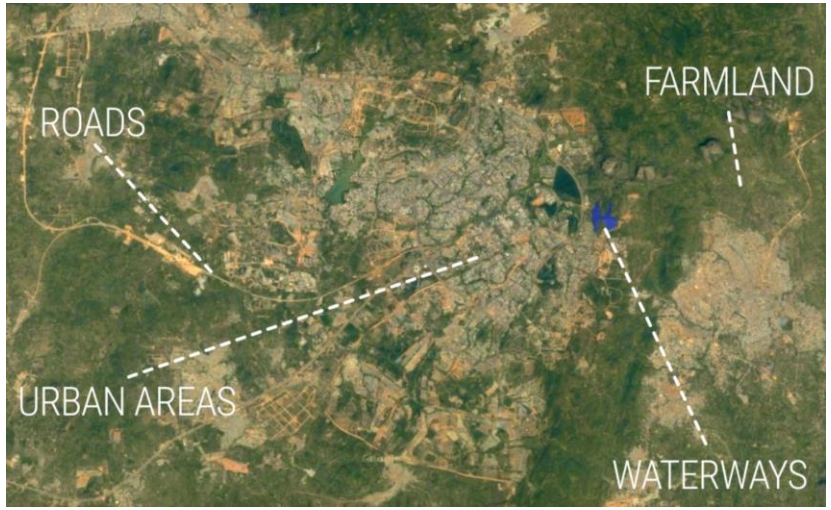
3D imaging
MRI, CT

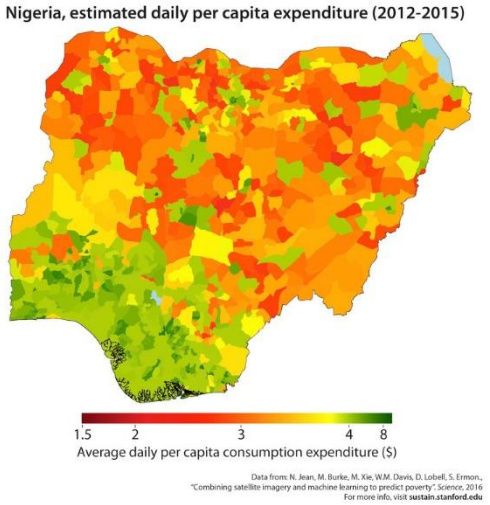
Image guided surgery
Grimson et al., MIT

Computer vision for the mass



Counting cells





Predicting poverty

Current state of the art

- Many of these are less than 5 years old
- Very active and exciting research area!
- To learn more about vision applications and companies
  - David Lowe maintains an excellent overview of vision companies
    - http://www.cs.ubc.ca/spider/lowe/vision.html

Course Overview

- ECE 4554 and ECE 5554
  - Tuesday and Thursday 3:30 pm to 4:45 pm
  - Surge Space Building 104C
- Office hours (Jia-Bin)
  - Friday 11 – 12 AM, 440 Whittemore Hall
- Office hours (Yuliang)
  - TBD
- Course webpage: http://bit.ly/vt-computer-vision-fall-2017
- Piazza discussion forum: http://piazza.com/vt/fall2017/ece5554ece4554/home
- HW/Project submission: https://canvas.vt.edu/

Grades

- Homework assignments (70%)
  - Five homework assignments: 100 points + graduate credit opportunities
  - ECE 4554: graded out of 525 points
  - ECE 5554: graded out of 600 points
  - Submission via https://canvas.vt.edu
- Final project (25%)
  - Proposal, project report webpage, and poster presentation
  - Work in a team of 2-4 students
- Attendance (5%)
  - Send TA and me a note if you cannot attend a class
- Late policy
  - Up to **five** free late days. After that, a penalty of 20% per day.

Graduate credits in homework assignments

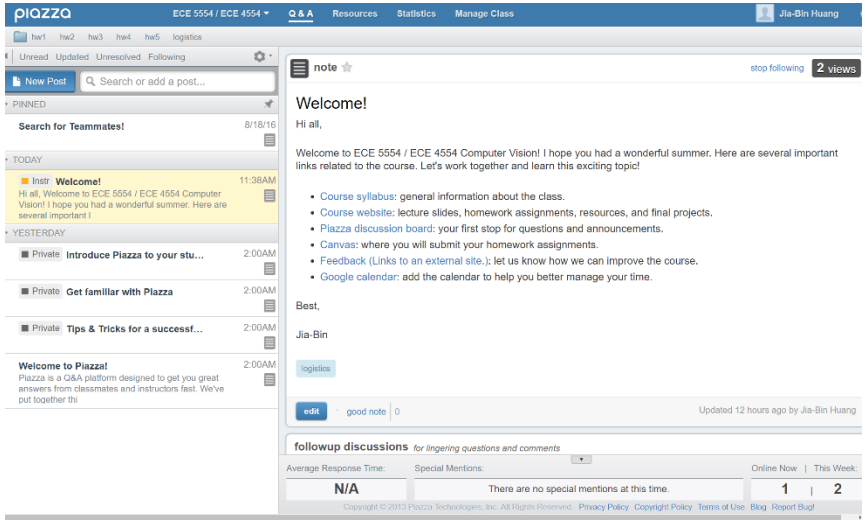
- Five homework assignments
- Each assignment: 100 points
- Graduate credits: 20 – 40 points for each assignment
  - Your choice
- ECE 4554: graded out of 525 points
  - On average 5 points per assignment
- ECE 5554: graded out of 600 points
  - On average 20 points

Academic Integrity

- Can discuss HW with peers, but don’t copy and/or share code
- Carefully document any sources within HW hand-in
- Do not use code from Internet unless you have permission
  - If you’re not sure, ask
- Do not use your published work as your final project

Getting help outside of class

Discussion Board:
http://piazza.com/vt/fall2017/ece5554ece4554/



Readings/Textbook:

http://szeliski.org/Book/

Lecture notes: will be posted online

TA: Yuliang Zou (ylzou@vt.edu)

Use Office Hours / After class

Office Hours

Source: [PhD Comics Movie 2]

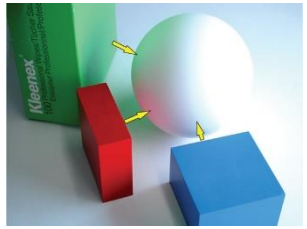
What to expect from this course

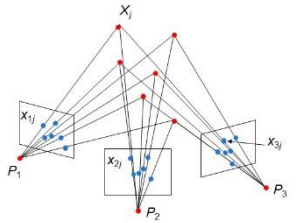
- Broad coverage
  - geometry, image processing, recognition, multiview, video
  - Focus is on algorithms, rather than specific systems.
- Background to delve deeper into any computer vision-related topic
- Practical experience
- Lots of work, tough material, fast pace, but hopefully lots of learning too!

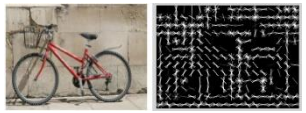
Other related courses at Virginia Tech

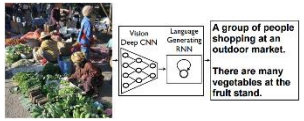
- Introductory courses:
  - Introduction to Machine Learning
  - Introduction to Artificial Intelligence
  - Computer Graphics
- Advanced courses:
  - Deep Learning
  - Probabilistic Graphical Models and Large-Scale Learning
  - Advanced Computer Vision
- Fundamentals:
  - ECE 5734 Convex Optimization
  - STAT 5444 Bayesian Statistics
  - STAT 4714 Prob and Stat for EE

Course Topics









- Interpreting Intensities
  - What determines the brightness and color of a pixel?
  - How can we use image filters to extract meaningful information from the image?
- Correspondence and Alignment
  - How can we find corresponding points in objects or scenes?
  - How can we estimate the transformation between them?
- Perspective and 3D Geometry
  - How can we map between the 3D world and the 2D image?
  - How can we recover 3D coordinates from images or video?
- Grouping and Segmentation
  - How can we group pixels into meaningful regions?
- Categorization and Object Recognition
  - How can we represent images and categorize them?
  - How can we recognize categories of objects?
- Advanced Topics
  - Action recognition, 3D scenes and context, CNNs, …

Prerequisites

- Linear algebra, basic calculus, and probability
  - Linear algebra review: http://cs229.stanford.edu/section/cs229-linalg.pdf

- Experience with image processing or MATLAB will help but is not necessary
  - Go through MATLAB Intro
  - Attend the MATLAB Tutorial section by Yuliang

Goals and Expectations

- My goal:
  - maximize the learning effectiveness of your time

- What I expect from you
  - Attend and participate, when possible

  - Start assignments well before deadline

  - Tell me what's working and suggest improvements
    Anonymous feedback form

Things to remember

- Computer vision is hard
- Lots of exciting and useful applications
- To-Do
  - Sign up piazza discussion board
  - Read course syllabus
  - Check out MATLAB Tutorials
  - Review Linear Algebra
- Next class: Light, shading, and color
- Questions?