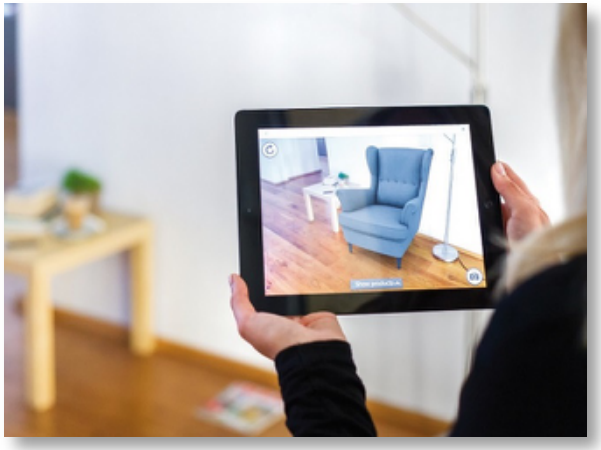
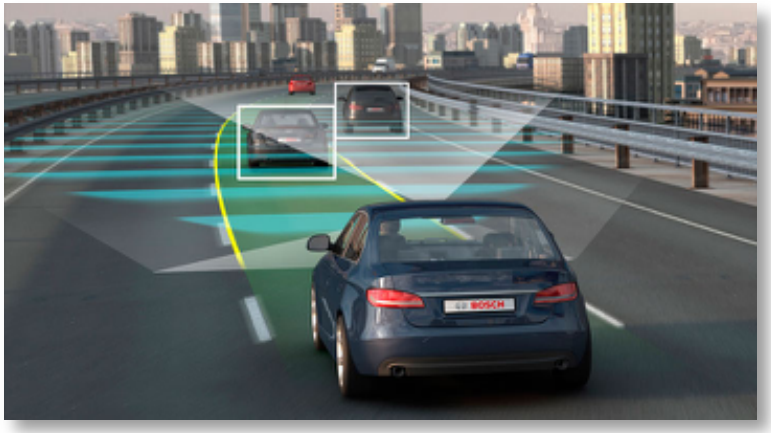
CS497P/597P: Computer Vision

Instructor: Scott Wehrwein





About Me

Scott Wehrwein

scott.wehrwein@wwu.edu

About Me

Research interests:

- **Computer vision** and graphics
- Computational photography and videography
- Photo and video enhancement
- Augmented reality

Today

1. What is computer vision?

2. Course overview

Today

- Readings
  - Szeliski, Chapter 1 (Introduction)

Every image tells a story

- Goal of computer vision: perceive the “story” behind the picture
- Compute properties of the world
  - 3D shape
  - Names of people or objects
  - What happened?

The goal of computer vision

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

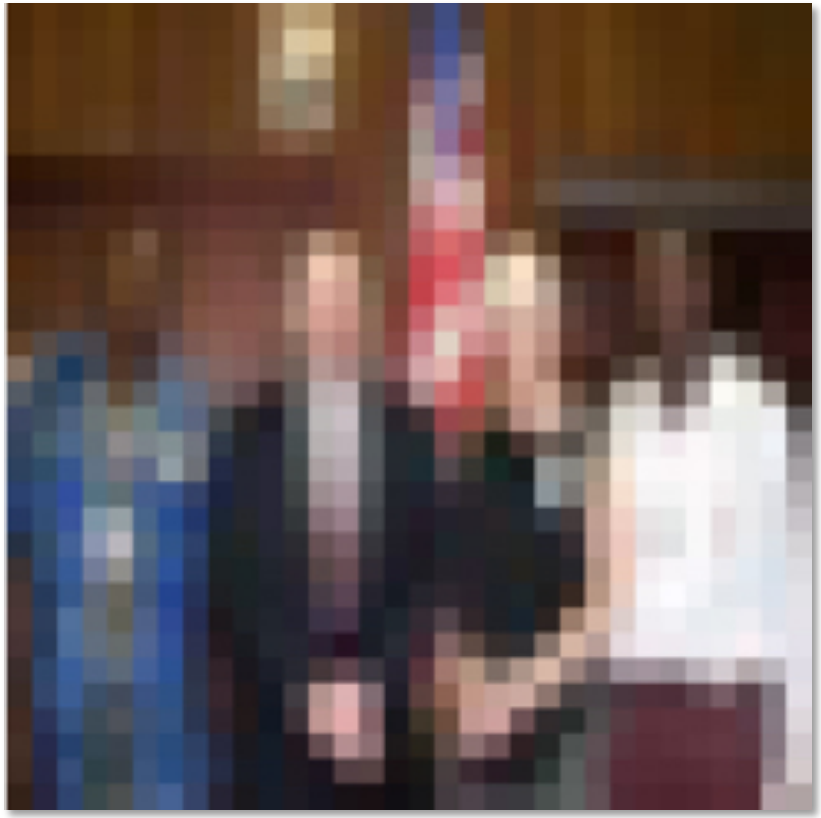
Can the computer match human perception?

- Yes and no (mainly no)
  - computers can be better at “easy” things
  - humans are much better at “hard” things
- But huge progress has been made
  - Accelerating in the last 4 years due to deep learning
  - What is considered “hard” keeps changing

Human perception has its shortcomings

Sinha and Poggio, *Nature*, 1996

But humans can tell a lot about a scene from a little information…



Source: “80 million tiny images” by Torralba, et al.

xkcd

9/24/2014

Introducing: Flickr PARK or BIRD

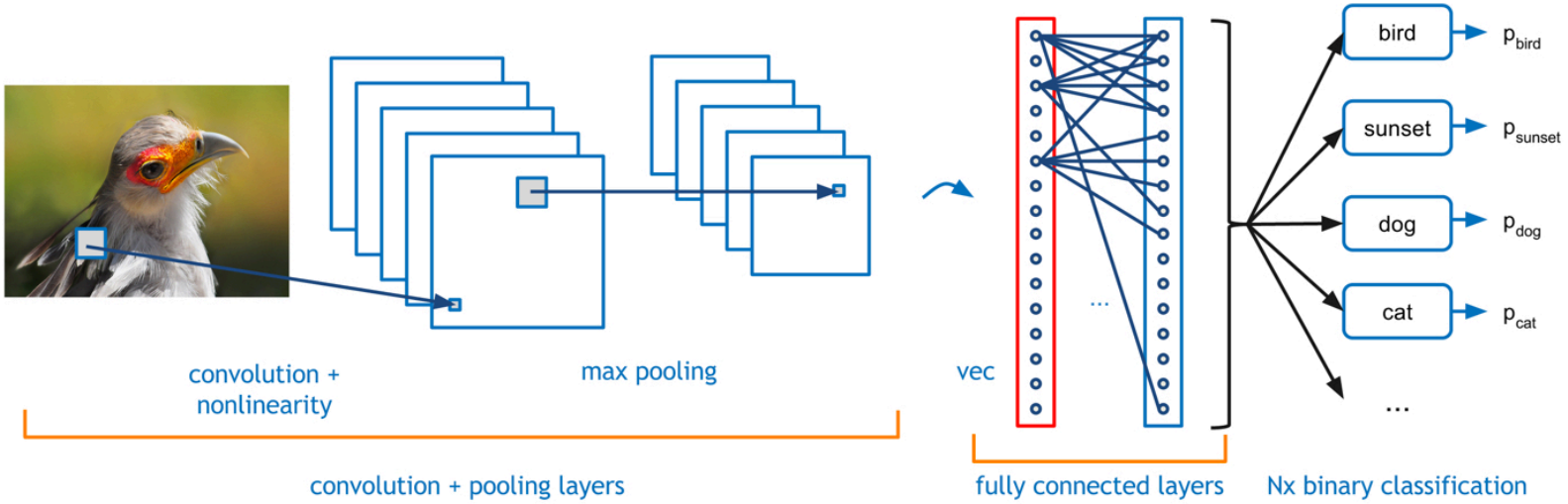
flickr

10/20/2014

Zion National Park Utah by Les Haines (cc) BY

OR

Secretary Bird by Bill Gracey (cc) BY-NC-ND



Why is computer vision difficult?

Viewpoint variation

Illumination

Scale

Why is computer vision difficult?

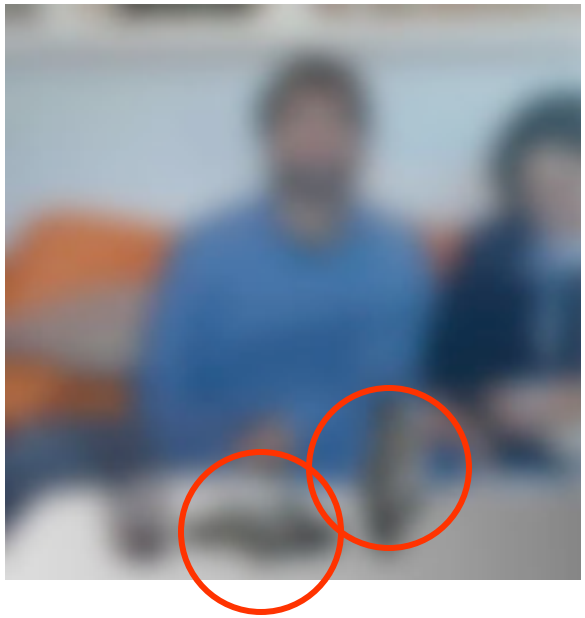
Intra-class variation

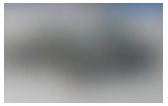
Motion (Source: S. Lazebnik)

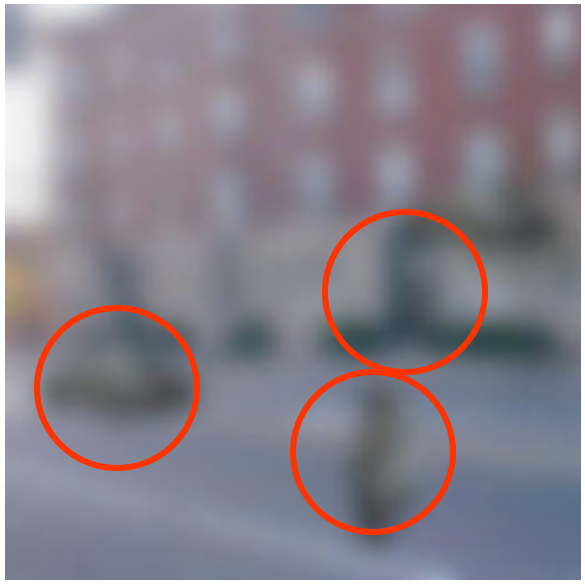
Background clutter

Occlusion

Challenges: local ambiguity







slide credit: Fei-Fei, Fergus & Torralba

But there are lots of cues we can exploit…

Source: S. Lazebnik

The state of Computer Vision and AI: we are really, really far.

Oct 22, 2012

The picture above is funny.

But for me it is also one of those examples that make me sad about the outlook for AI and for Computer Vision. What would it take for a computer to understand this image as you or I do? I challenge you to think explicitly of all the pieces of knowledge that have to fall in place for it to make sense. Here is my short attempt:

- You recognize it is an image of a bunch of people and you understand they are in a hallway
- You recognize that there are 3 mirrors in the scene so some of those people are "fake" replicas from different viewpoints.
- You recognize Obama from the few pixels that make up his face. It helps that he is in his suit and that he is surrounded by other people with suits.
- You recognize that there's a person standing on a scale, even though the scale occupies only very few white pixels that blend with the background. But, you've used the person's pose and knowledge of how people interact with objects to figure it out.
- You recognize that Obama has his foot positioned just slightly on top of the scale. Notice the language I'm using: It is in terms of the 3D structure of the scene, not the position of the leg in the 2D coordinate system of the image.
- You know how physics works: Obama is leaning in on the scale, which applies a force on it. Scale measures force that is applied on it, that's how it works => it will over-estimate the weight of the person standing on it.
- The person measuring his weight is not aware of Obama doing this. You derive this because you know his pose, you understand that the field of view of a person is finite, and you understand that he is not very likely to sense the slight push of Obama's foot.
- You understand that people are self-conscious about their weight. You also understand that he is reading off the scale measurement, and that shortly the over-estimated weight will confuse him because it will probably be much higher than what he expects. In other words, you reason about implications of the events that are about to unfold seconds after this photo was taken, and especially about the thoughts and how they will develop inside people's heads. You also reason about what pieces of information are available to people.
- There are people in the back who find the person's imminent confusion funny. In other words you are reasoning about state of mind of people, and their view of the state of mind of another person. That's getting frighteningly meta.
- Finally, the fact that the perpetrator here is the president makes it maybe even a little more funnier. You understand what actions are more or less likely to be undertaken by different people based on their status and identity.

The state of Computer Vision and AI: we are really, really far.

Oct 22, 2012

The picture above is funny.

But for me it is also one of those examples that make me sad about the outlook for AI and for Computer Vision. What would it take for a computer to understand this image as you or I do? I challenge you to think explicitly of all the pieces of knowledge that have to fall in place for it to make sense. Here is my short attempt:

- You recognize it is an image of a bunch of people and you understand they are in a hallway
- You recognize that there are 3 mirrors in the scene so some of those people are "fake" replicas from different viewpoints.
- You recognize Obama from the few pixels that make up his face. It helps that he is in his suit and that he is surrounded by other people with suits.
- You recognize that there's a person standing on a scale, even though the scale occupies only very few white pixels that blend with the background. But, you've used the person's pose and knowledge of how people interact with objects to figure it out.
- You recognize that Obama has his foot positioned just slightly on top of the scale. Notice the language I'm using: It is in terms of the 3D structure of the scene, not the position of the leg in the 2D coordinate system of the image.
- You know how physics works: Obama is leaning in on the scale, which applies a force on it. Scale measures force that is applied on it, that's how it works => it will over-estimate the weight of the person standing on it.
- The person measuring his weight is not aware of Obama doing this. You derive this because you know his pose, you understand that the field of view of a person is finite, and you understand that he is not very likely to sense the slight push of Obama's foot.
- You understand that people are self-conscious about their weight. You also understand that he is reading off the scale measurement, and that shortly the over-estimated weight will confuse him because it will probably be much higher than what he expects. In other words, you reason about implications of the events that are about to unfold seconds after this photo was taken, and especially about the thoughts and how they will develop inside people's heads. You also reason about what pieces of information are available to people.
- There are people in the back who find the person's imminent confusion funny. In other words you are reasoning about state of mind of people, and their view of the state of mind of another person. That's getting frighteningly meta.
- Finally, the fact that the perpetrator here is the president makes it maybe even a little more funnier. You understand what actions are more or less likely to be undertaken by different people based on their status and identity.

Bottom line

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

  - We often need to use prior knowledge about the structure of the world

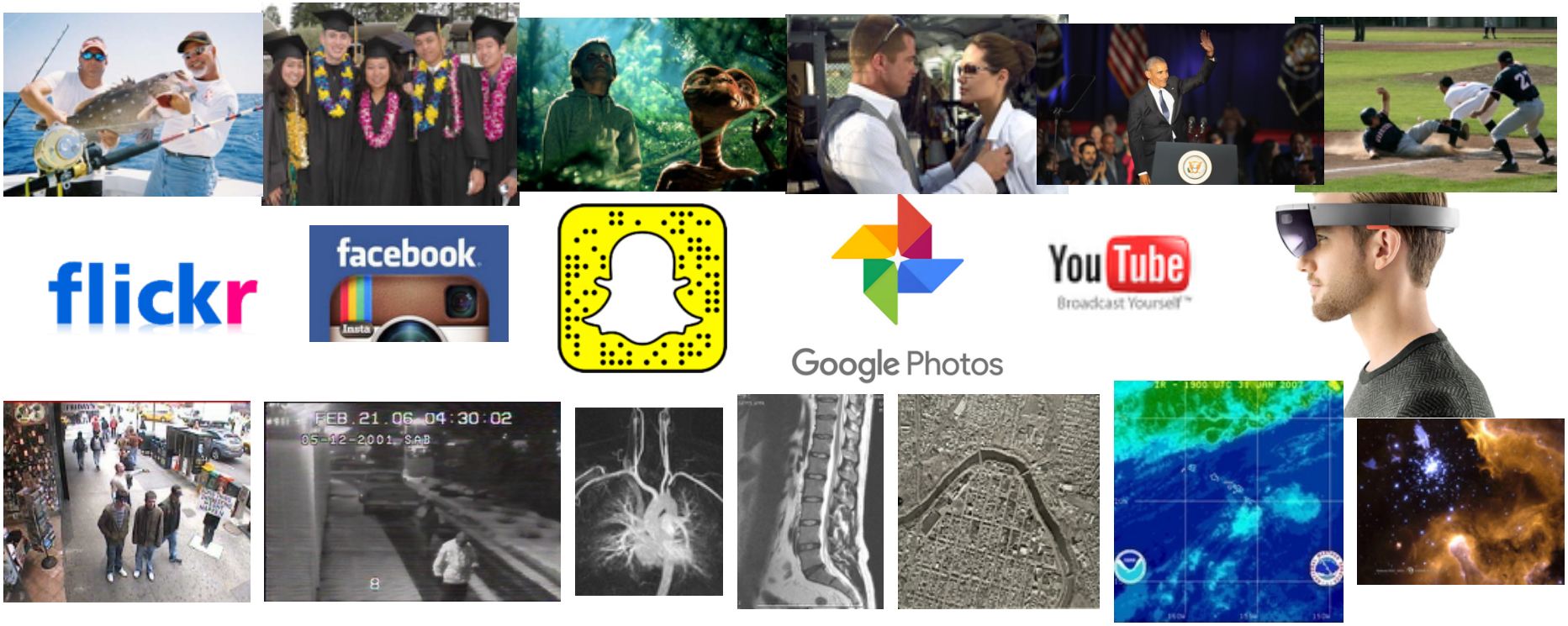
The goals of computer vision

https://www.youtube.com/watch?v=9MeaaCwBW28

The goals of computer vision

Why study computer vision?

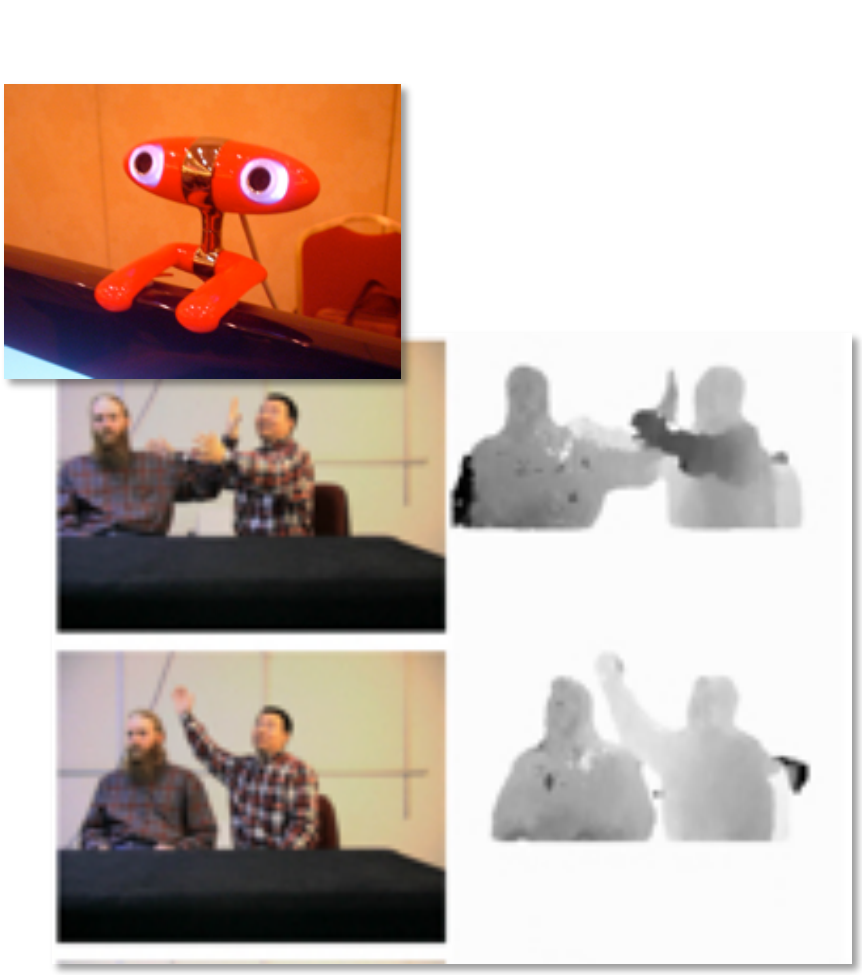
- Billions of images/videos captured per day

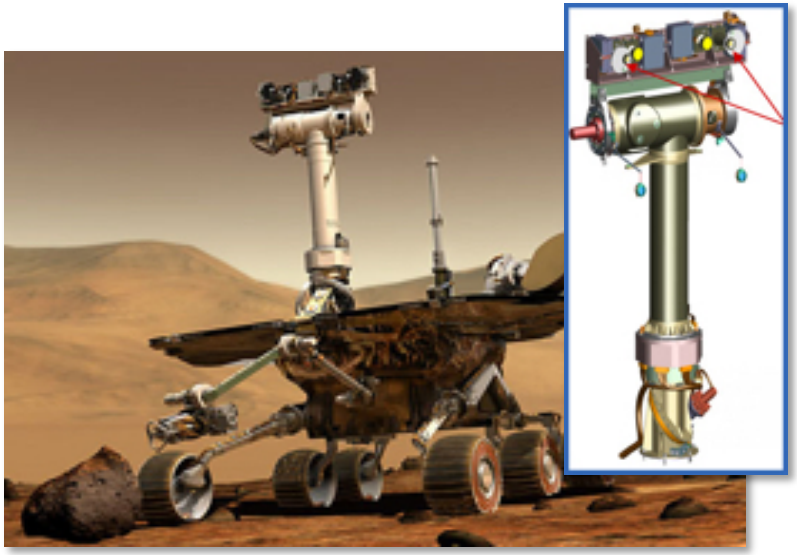


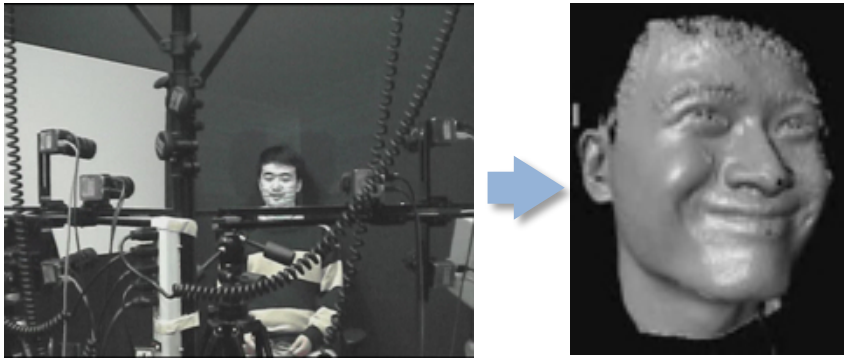
- Huge number of useful applications
- The next slides show the current state of the art

The goals of computer vision

- Compute the 3D shape of the world

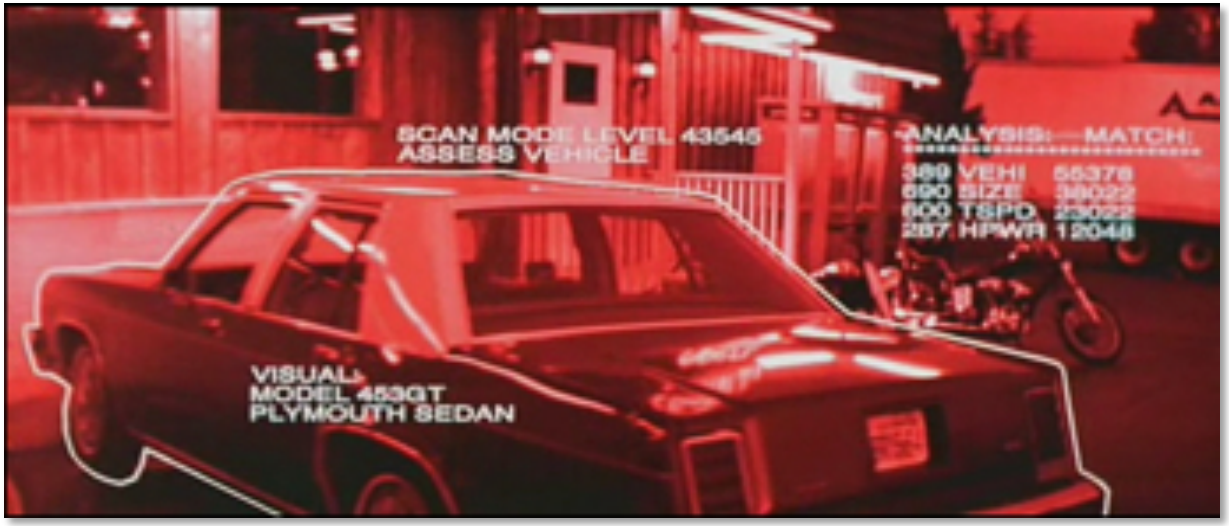






The goals of computer vision

- Recognize objects and people



Terminator 2, 1991


中华人民
和国万岁
世界
民大团结万岁

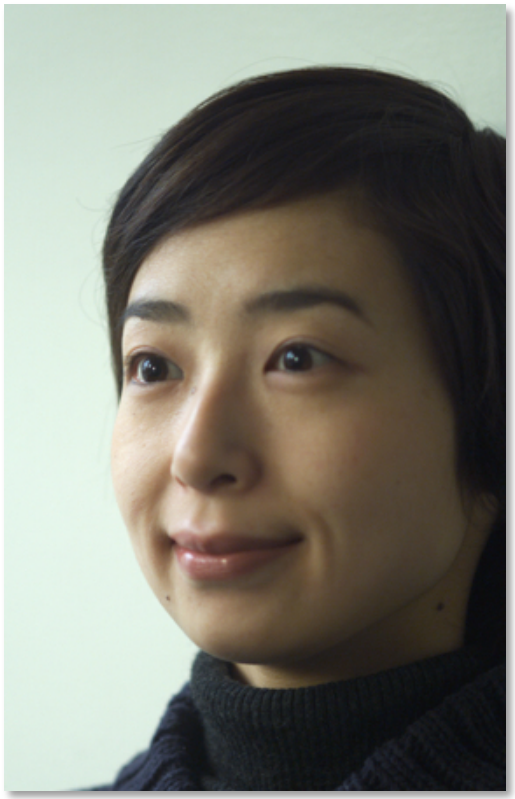

sky
building
flag
face
banner
wall
street lamp
bus
bus
cars
slide credit: Fei-Fei, Fergus & Torralba

The goals of computer vision

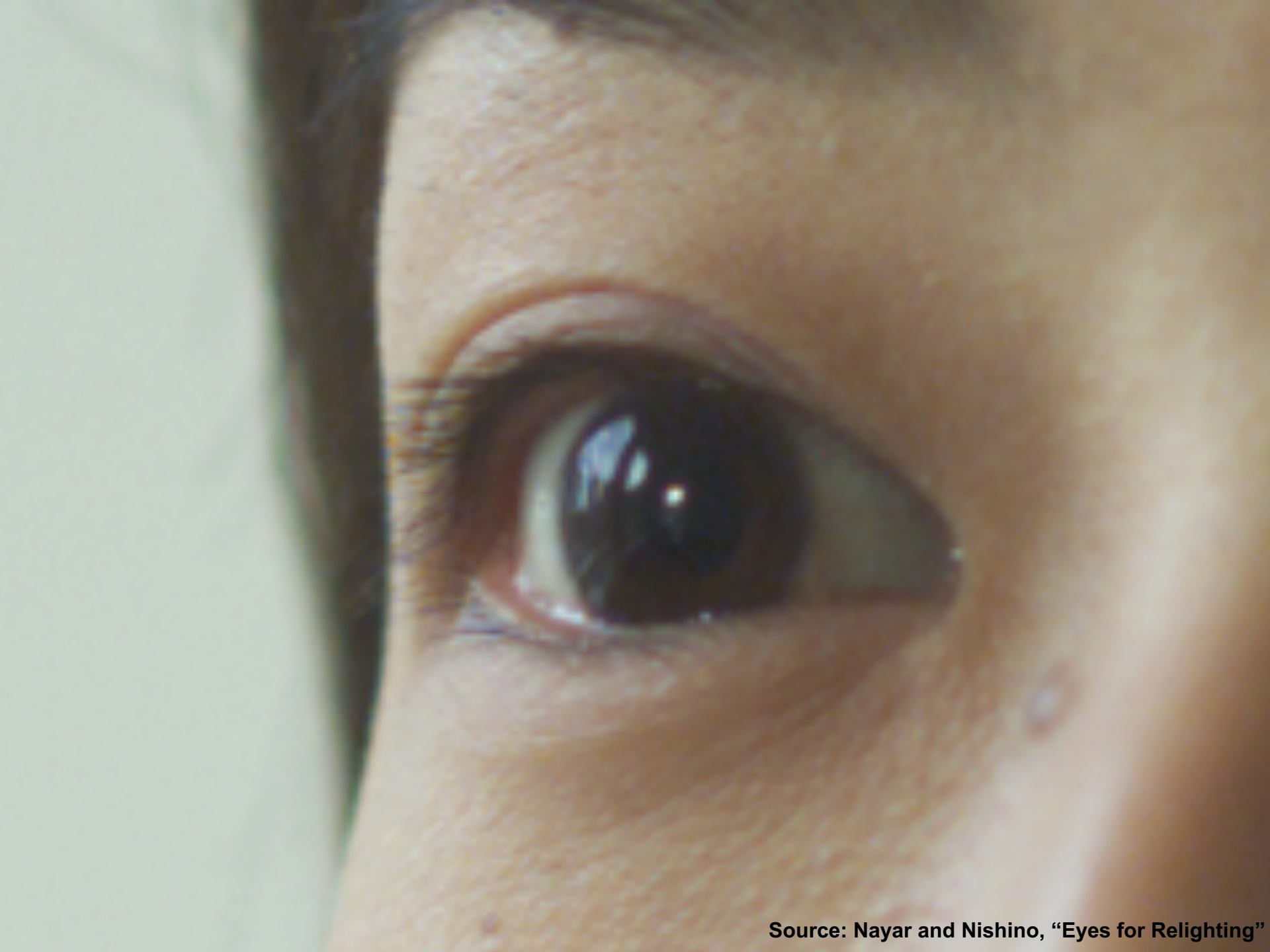
- “Enhance” images

The goal of computer vision

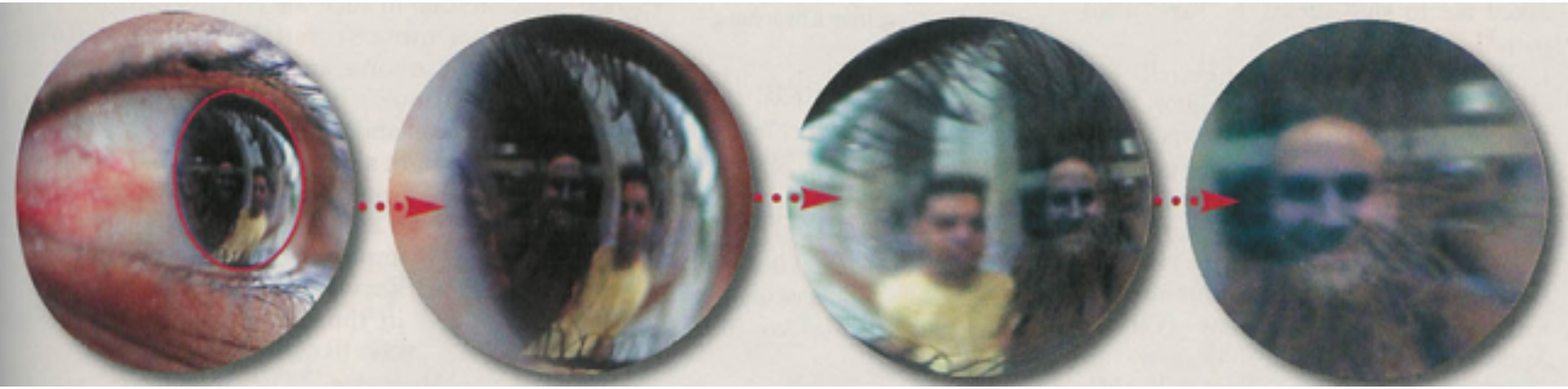
- Forensics



Source: Nayar and Nishino, “Eyes for Relighting”



Source: Nayar and Nishino, “Eyes for Relighting”



Source: Nayar and Nishino, “Eyes for Relighting”

Researchers warn peace sign photos could expose fingerprints

But the likelihood of anyone actually using images to recreate prints is pretty slim.

Jamie Rigg, @jmerigg
01.13.17 in Security

Comments

1721
Shares

Getty

The goals of computer vision

- Improve photos ("Computational Photography")

Super-resolution (source: 2d3)

Low-light photography
(credit: Hasinoff et al., SIGGRAPH ASIA 2016)

Depth of field on cell phone camera
(source: Google Research Blog)

Inpainting / image completion
(image credit: Hays and Efros)

Optical character recognition (OCR)

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

Digit recognition, AT&T labs (1990's)
http://yann.lecun.com/exdb/lenet/

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

Automatic check processing

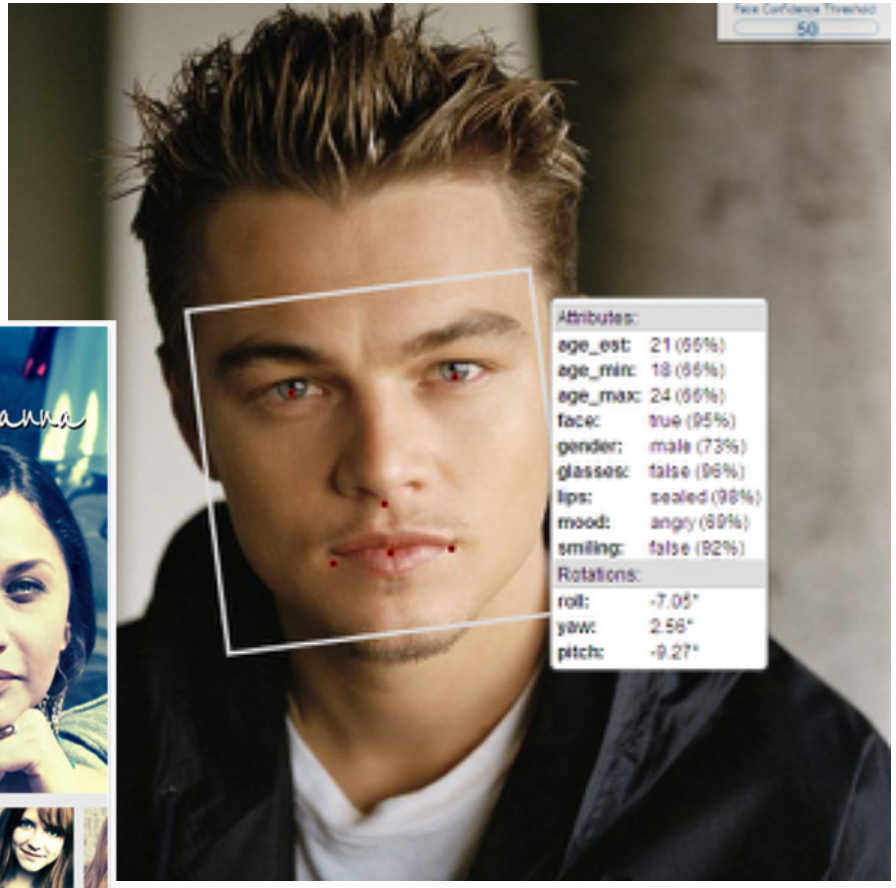
Sudoku grabber
http://sudokugrab.blogspot.com/

Face detection

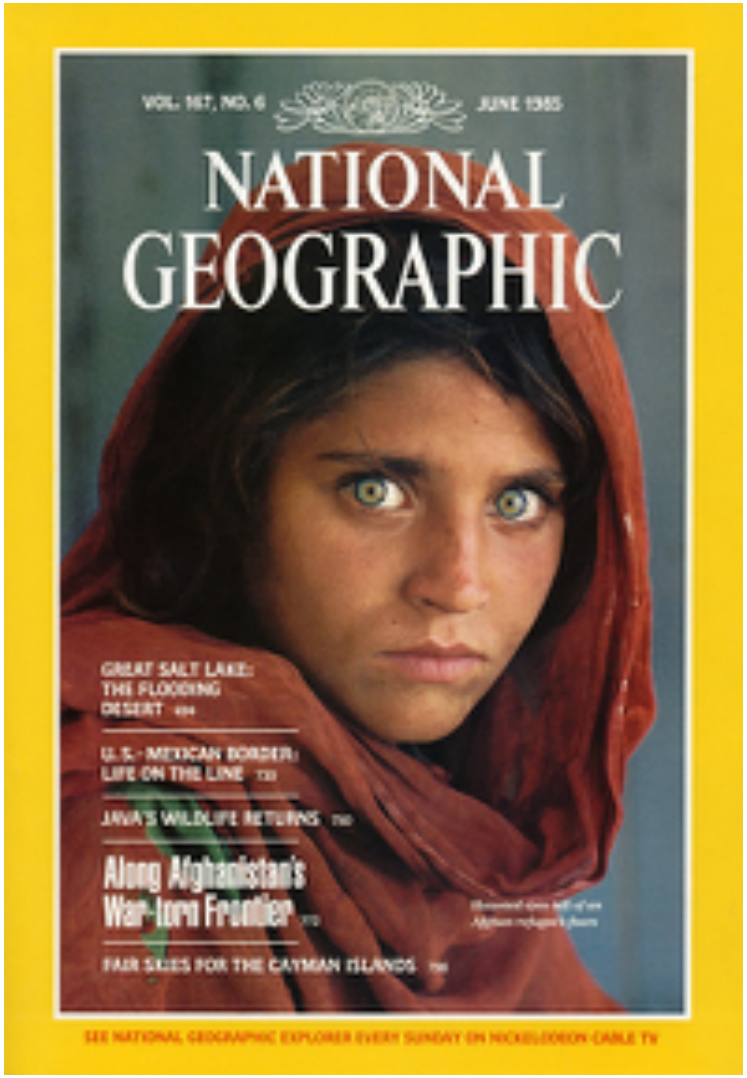
- Nearly all cameras detect faces in real time
  - (Why?)

Face Recognition

FACE++



Face recognition

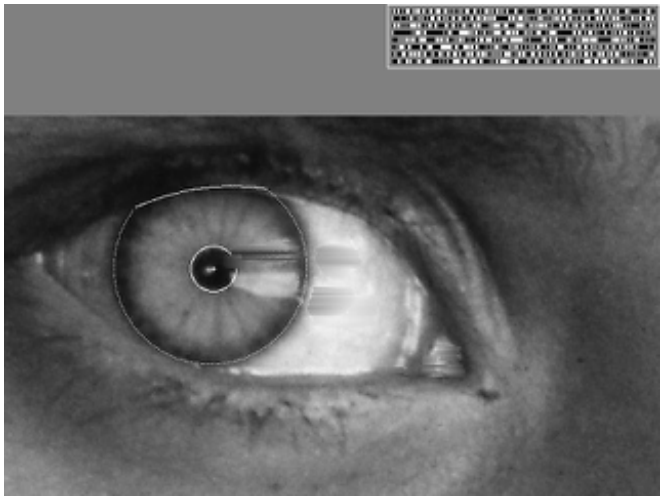


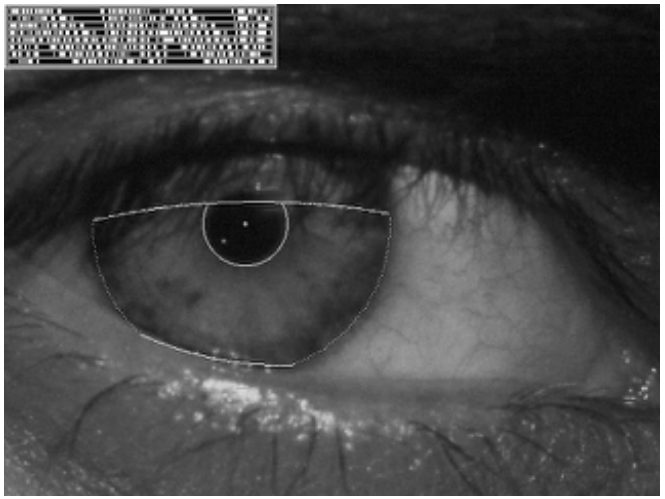
Who is she?

Source: S. Seitz

Vision-based biometrics

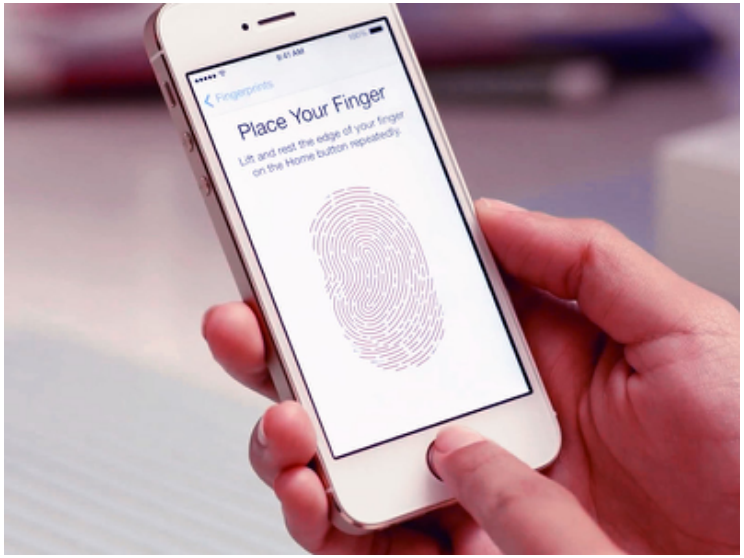
"*How the Afghan Girl was Identified by Her Iris Patterns*" Read the story





Source: S. Seitz

Login without a password



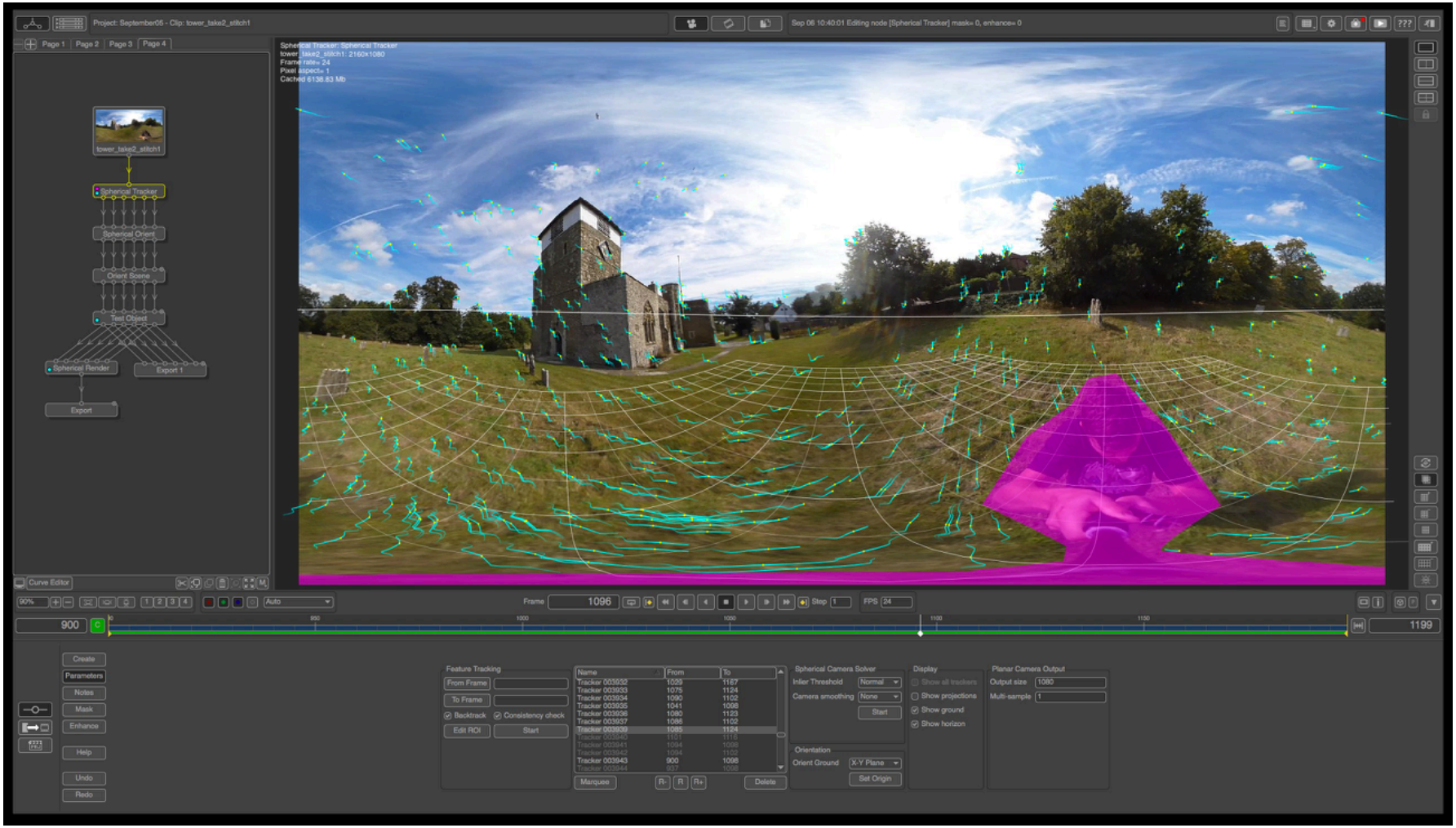
Fingerprint scanners on many new smartphones and other devices

Face unlock on Apple iPhone X
See also http://www.sensiblevision.com/

Bird Identification

Merlin Bird ID

Special effects: camera tracking



PFTrack

Special effects: shape capture

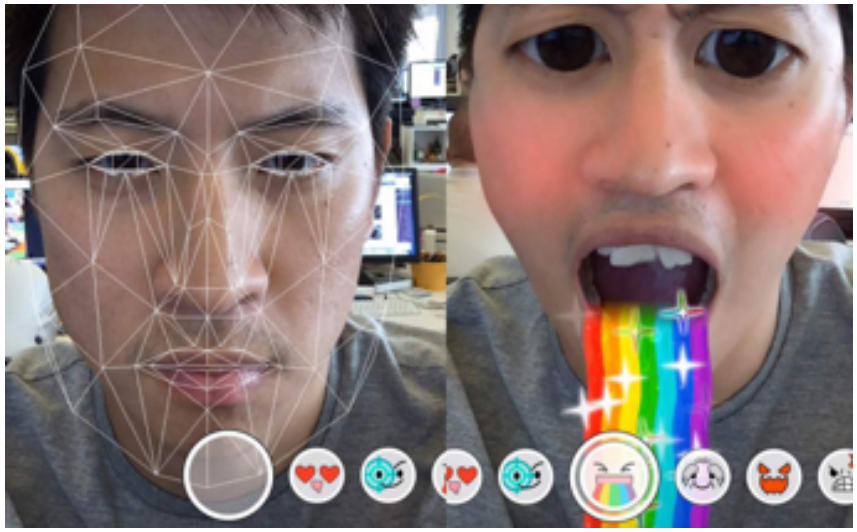
The Matrix movies, ESC Entertainment, XYZRGB, NRC

Source: S. Seitz

Special effects: motion capture

Pirates of the Carribean, Industrial Light and Magic

3D face tracking w/ consumer cameras



Snapchat Lenses

Face2Face system (Thies et al.)

Sports

Sportvision first down line
Nice [explanation](#) on www.howstuffworks.com

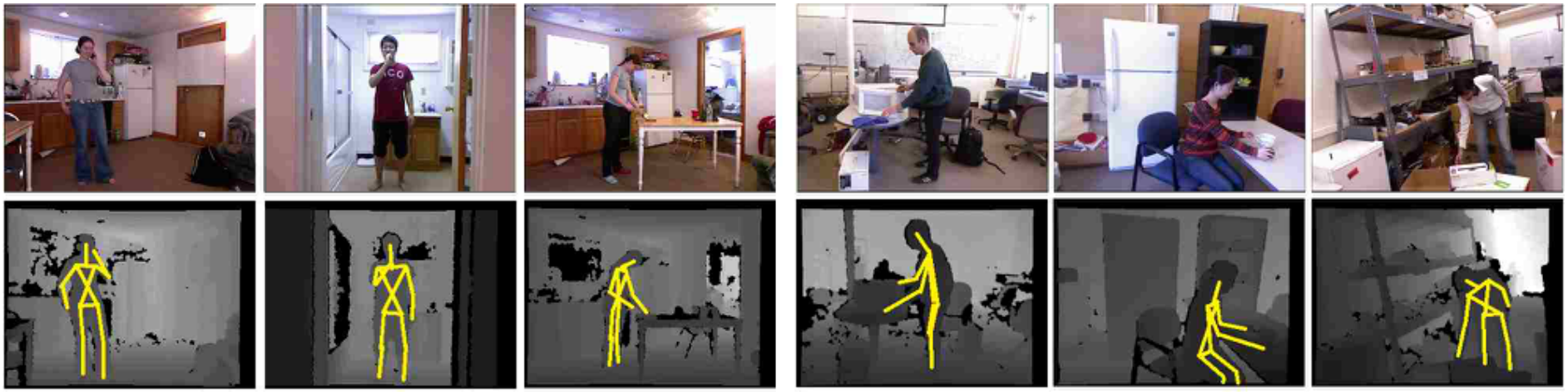
Source: S. Seitz

Vision-based interaction (and games)

Nintendo Wii has camera-based IR tracking built in. See Lee's work at CMU on clever tricks on using it to create a multi-touch display!

Assistive technologies

Kinect



RGB, depth, and Pose Estimation

Smart cars

- Mobileye
- Tesla Autopilot
- Safety features in many high-end cars

Self-driving cars

Waymo, Uber, and many others

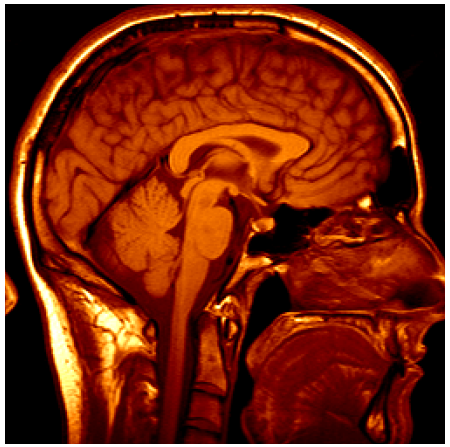
Robotics

NASA's Mars Curiosity Rover
https://en.wikipedia.org/wiki/Curiosity_(rover)

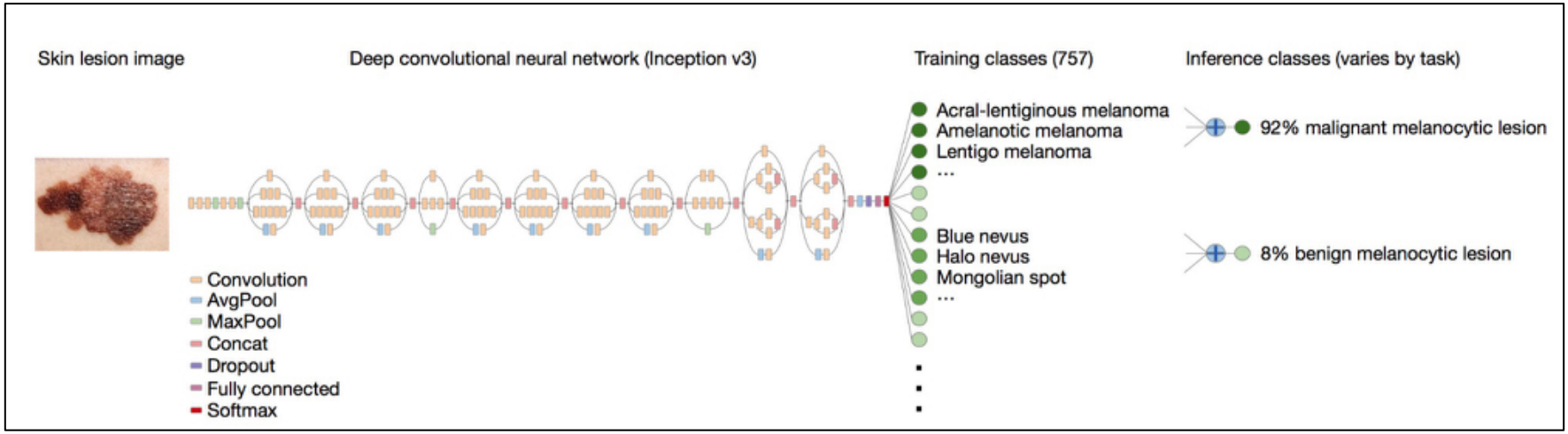
Amazon Picking Challenge
http://www.robocup2016.org/en/events/amazon-picking-challenge/

Amazon Prime Air

Medical imaging



3D imaging
(MRI, CT)



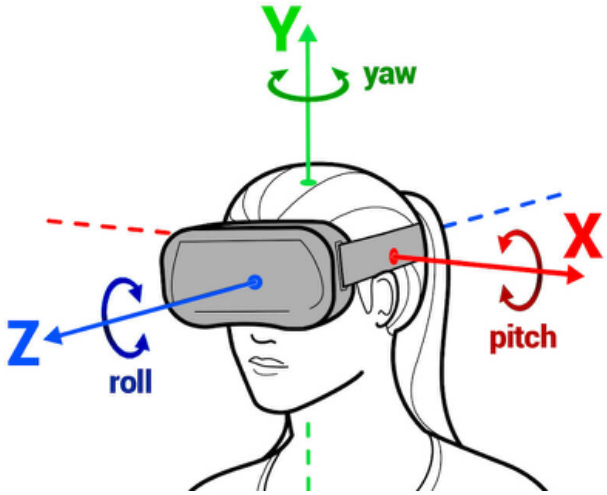
Skin cancer classification with deep learning
https://cs.stanford.edu/people/esteva/nature/

INVESTING 3/25/2014 @ 5:43PM | 70,399 views

Facebook Buys Oculus, Virtual Reality Gaming Startup, For $2 Billion

+ Comment Now + Follow Comments

Virtual & Augmented Reality



6DoF head tracking

Hand & body tracking

3D scene understanding

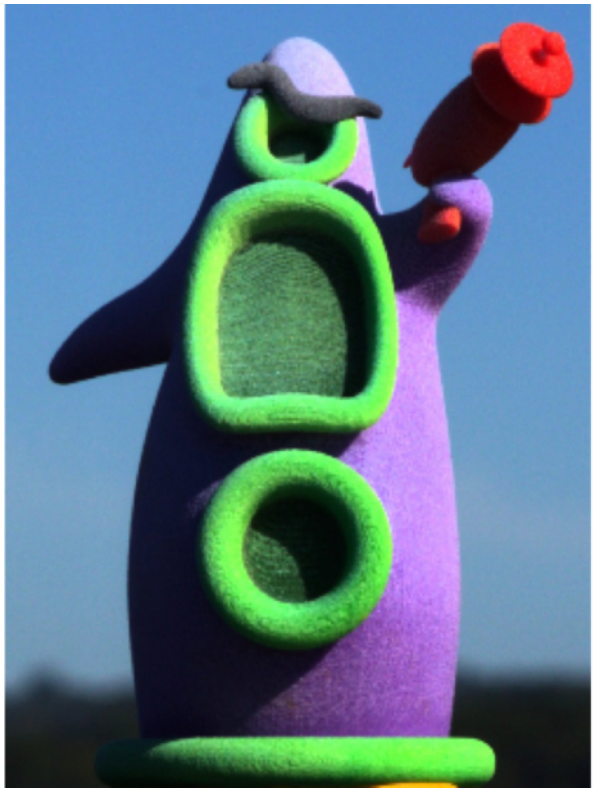
3D-360 video capture

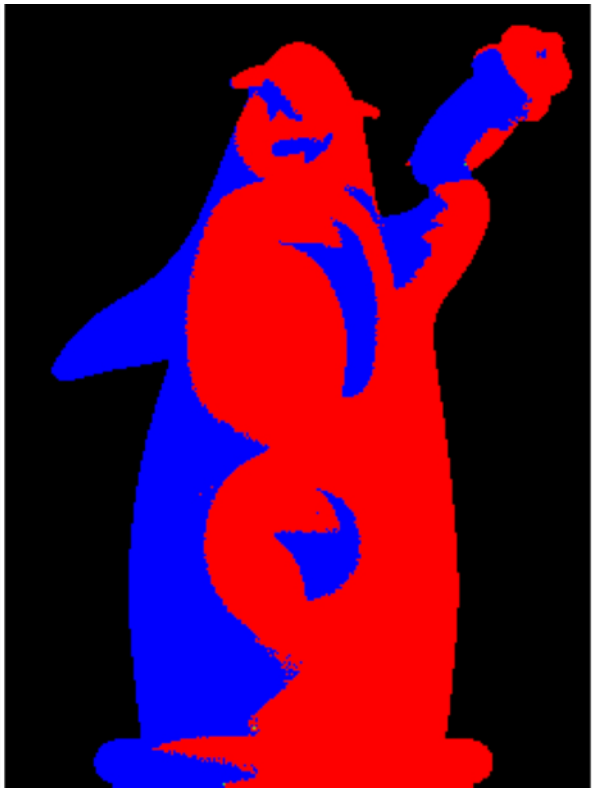
Current state of the art

- You just saw many examples of current systems.
  - Many of these are less than 5 years old
- This is a very active research area, and rapidly changing
  - Many new apps in the next 5 years
  - Deep learning powering many modern applications
- Many startups across a dizzying array of areas
  - VR/AR, deep learning, robotics, autonomous vehicles, medical imaging, construction, manufacturing, …

My Work: Video Segmentation

My Work: Illumination Estimation

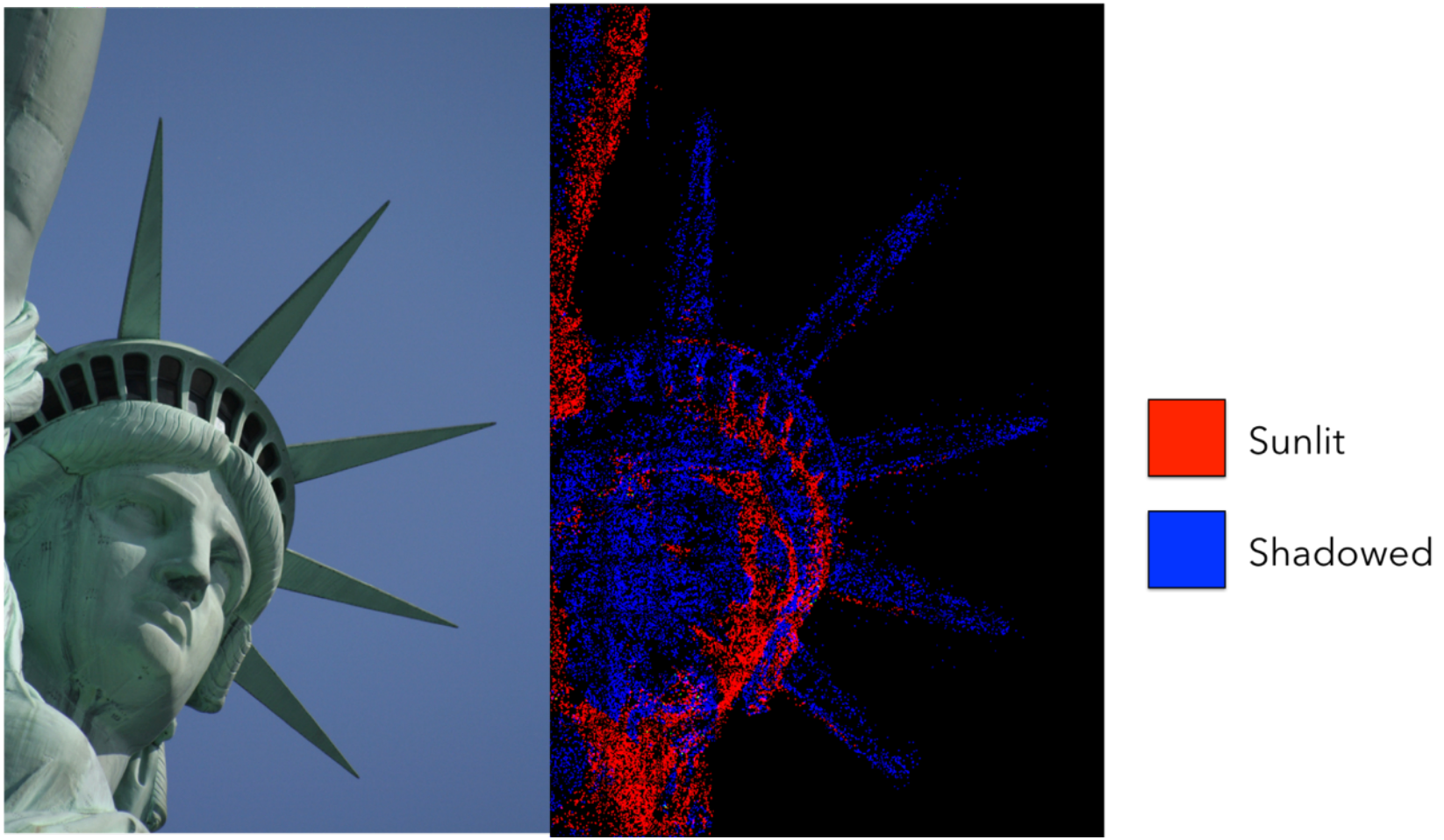


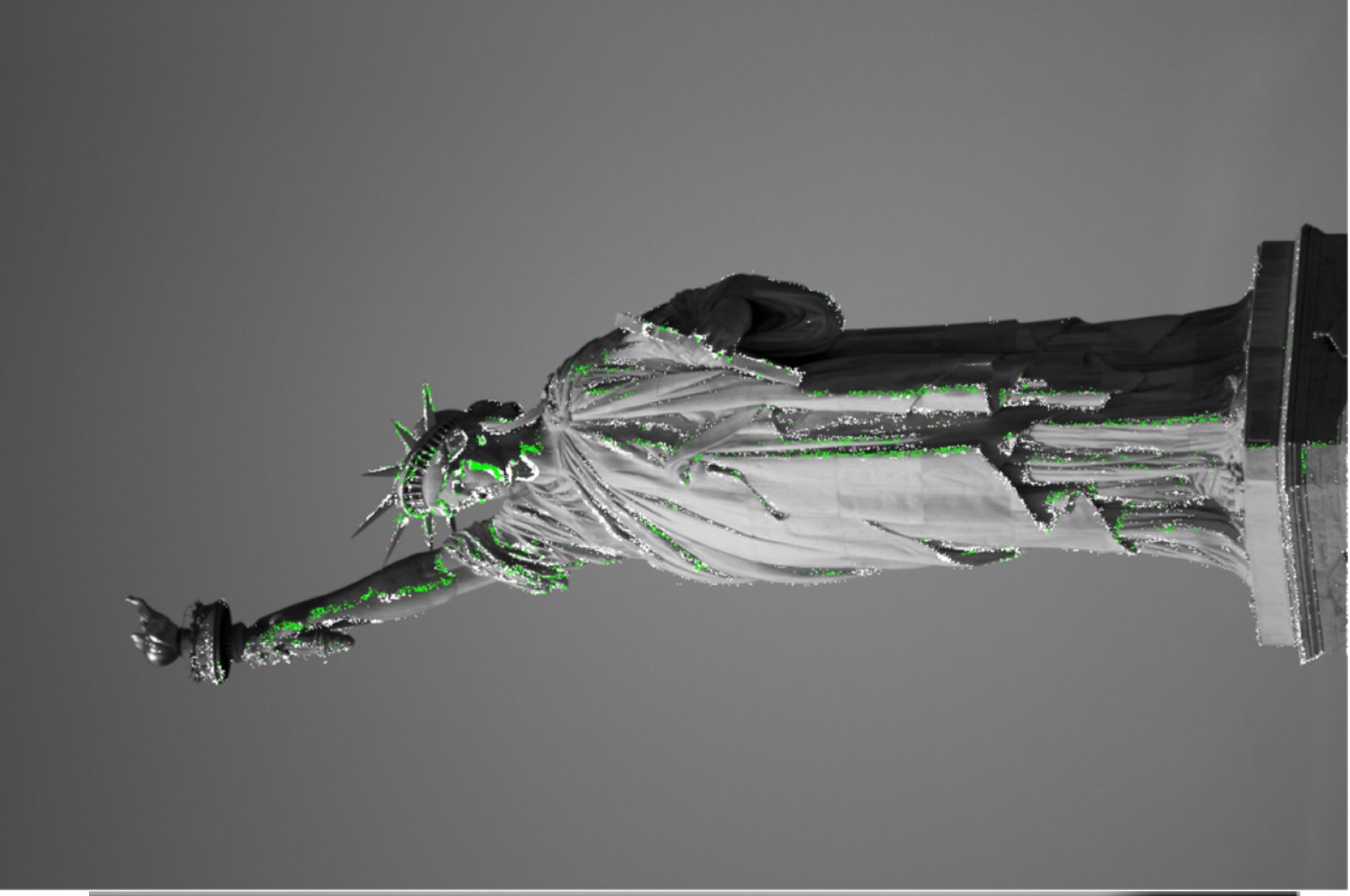


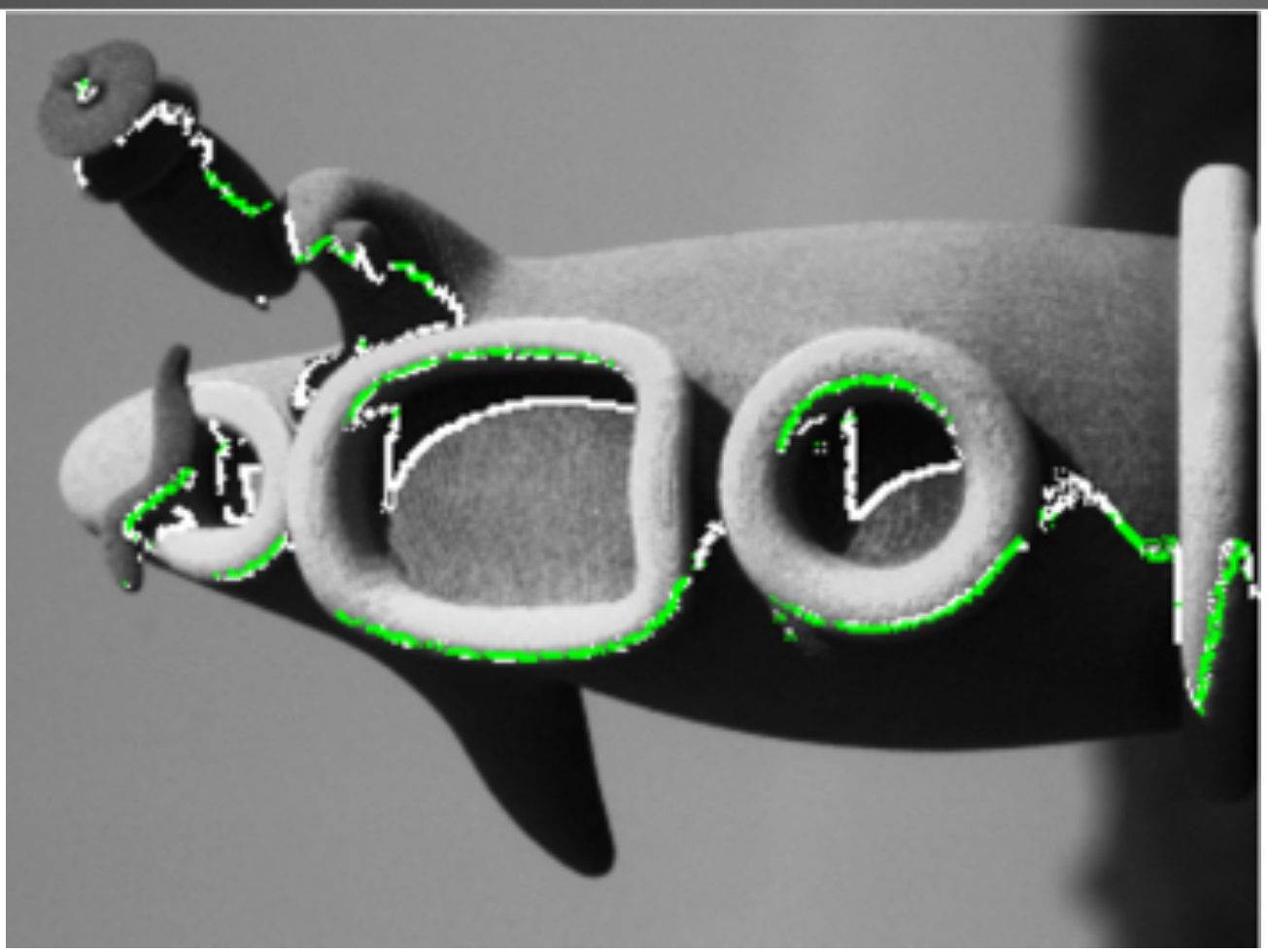
Sunlit

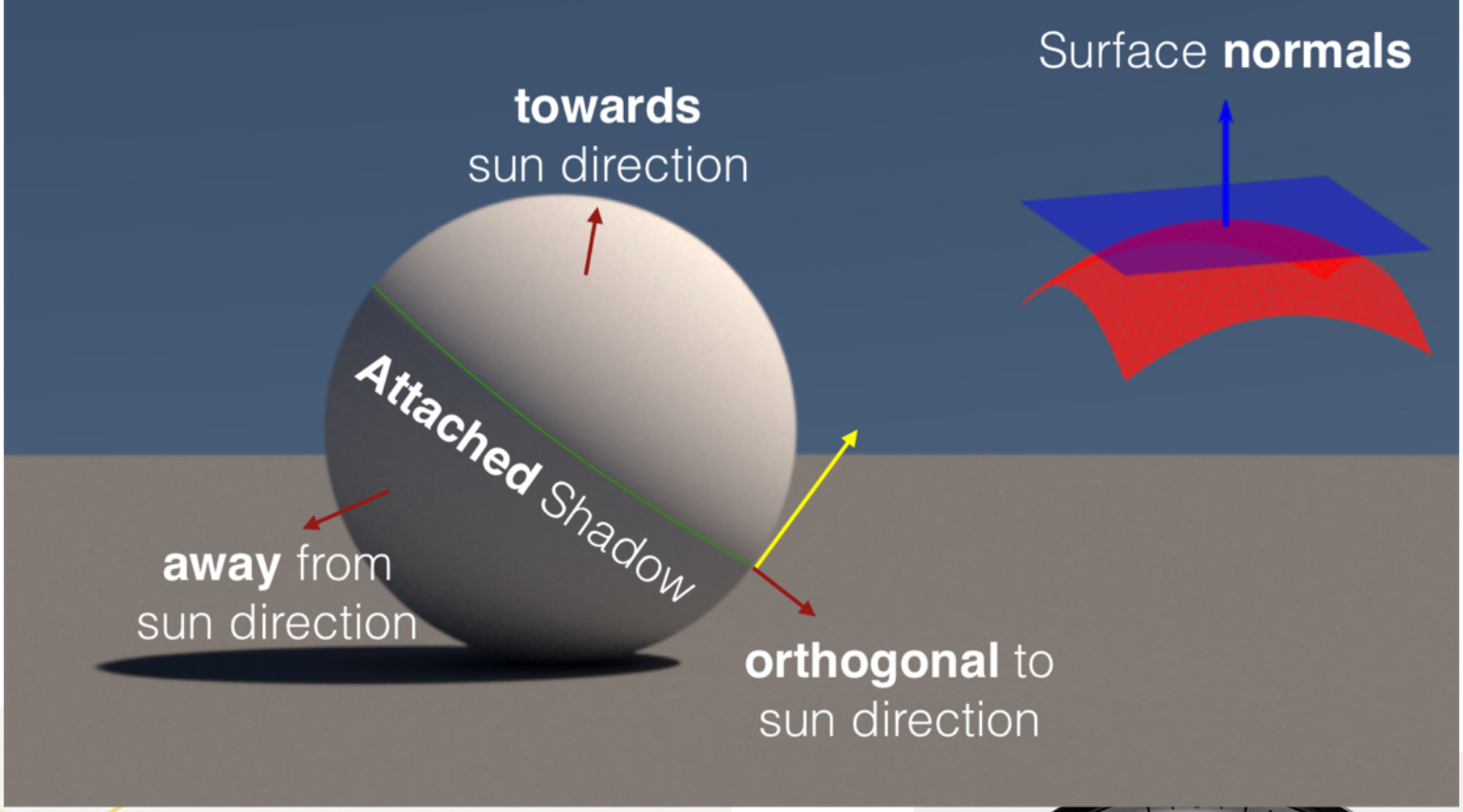
Shadowed

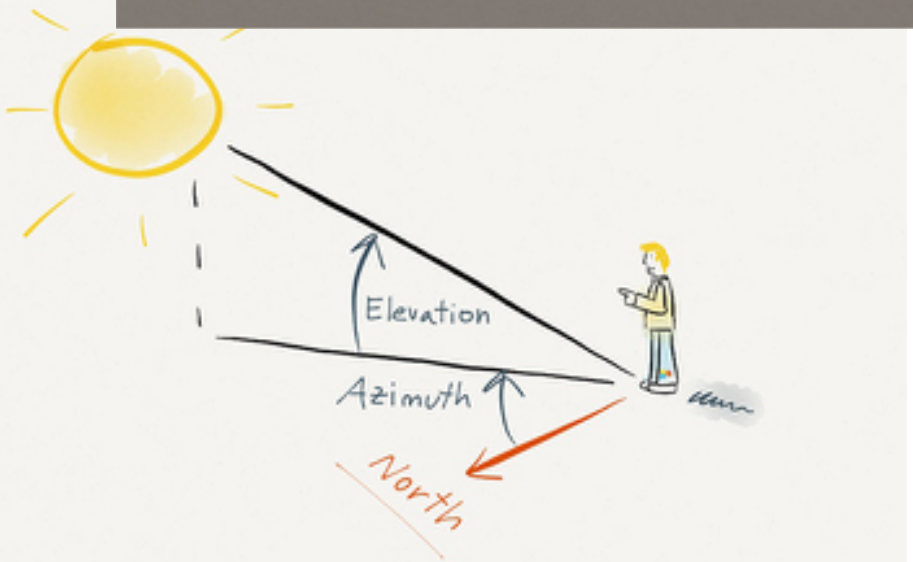
My Work: Illumination Estimation

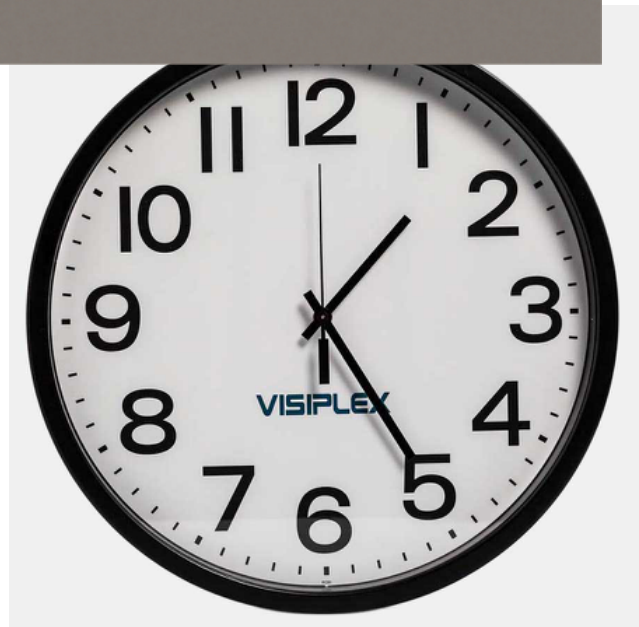






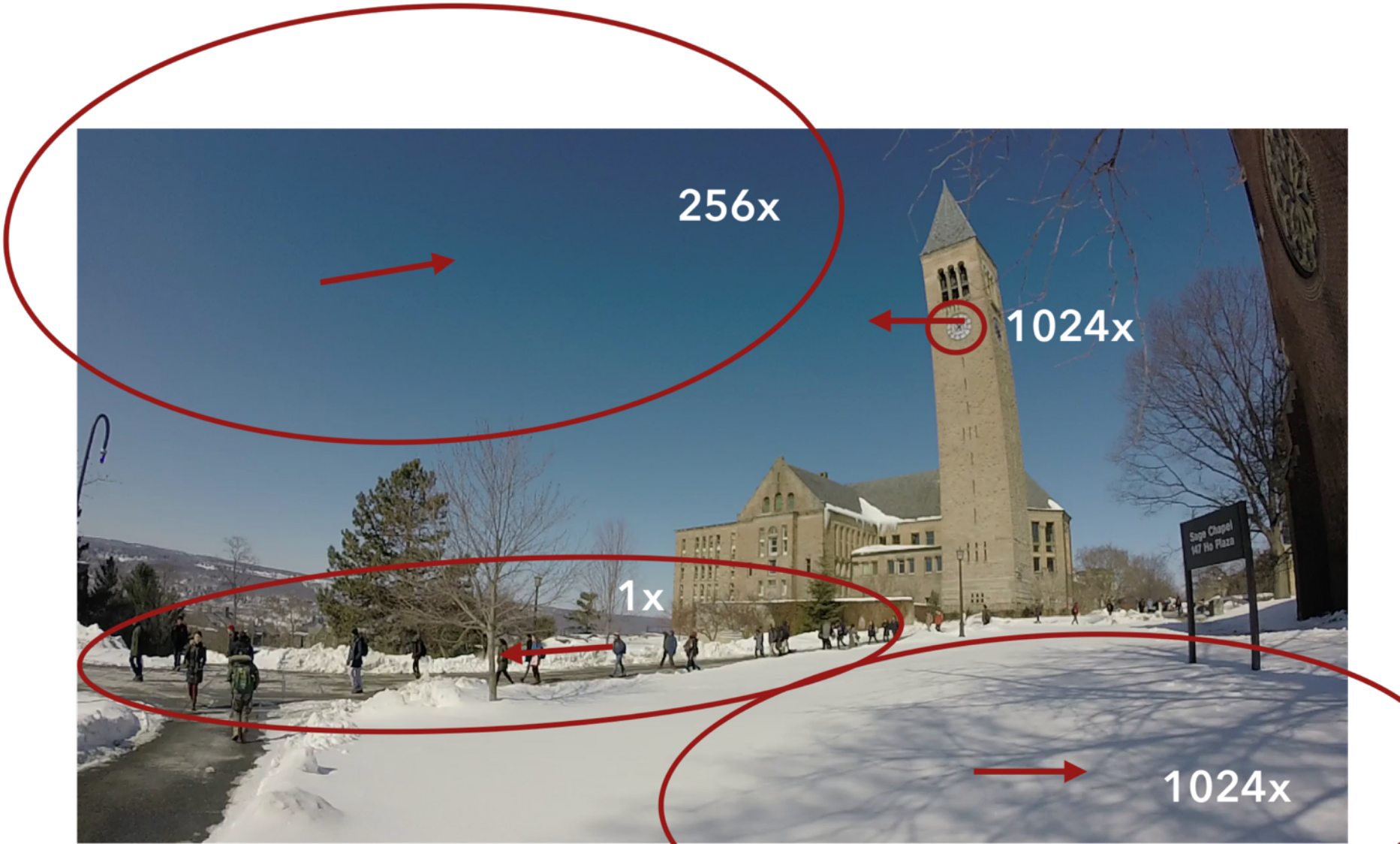

towards
sun direction
Surface **normals**
Attached Shadow
away from
sun direction
orthogonal to
sun direction


Elevation
Azimuth
North


VISIPLEX

My Work: Timelapse Visualization

My Work: Timelapse Visualization



My Work: Timelapse Visualization

My Work: Timelapse Visualization

My Work: Timelapse Visualization

CVPR Attendance

xkcd

9/24/2014

Introducing: Flickr PARK or BIRD

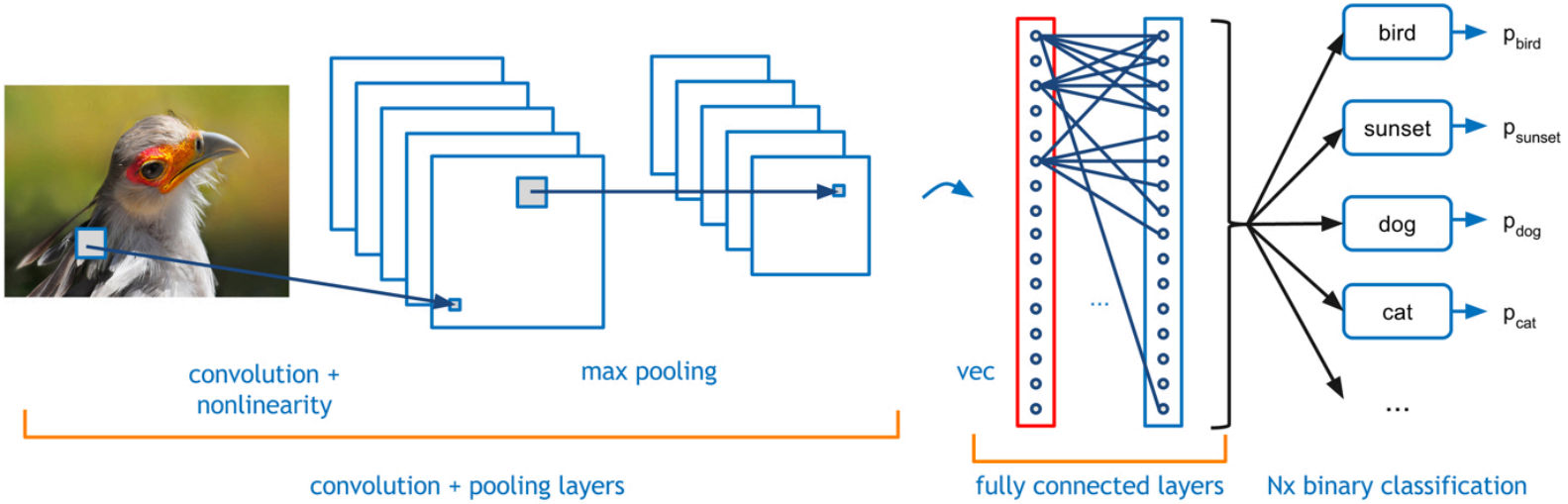
flickr

10/20/2014

Zion National Park Utah by Les Haines (cc) BY

OR

Secretary Bird by Bill Gracey (cc) BY-NC-ND



Why is computer vision difficult?

Viewpoint variation

Illumination

Scale

Why is computer vision difficult?

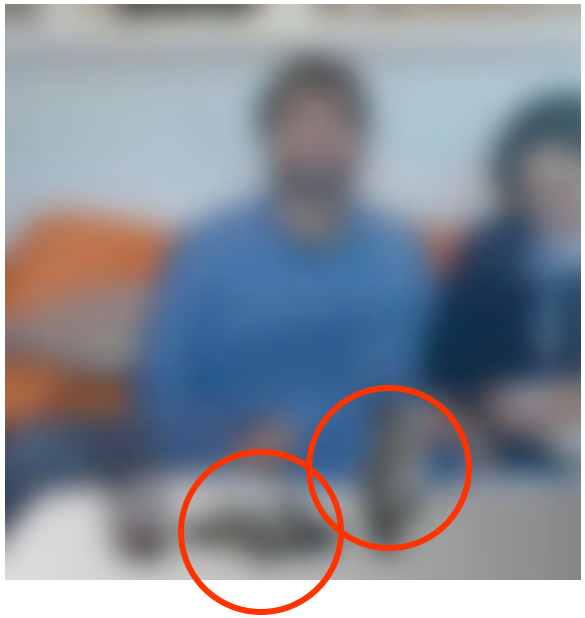
Intra-class variation

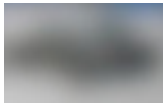
Motion (Source: S. Lazebnik)

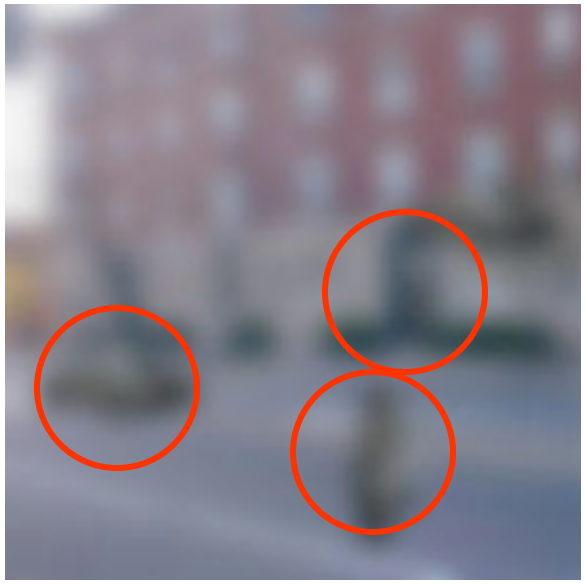
Background clutter

Occlusion

Challenges: local ambiguity







slide credit: Fei-Fei, Fergus & Torralba

But there are lots of cues we can exploit…

Source: S. Lazebnik

Bottom line

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

  - We often need to use prior knowledge about the structure of the world

The state of Computer Vision and AI: we are really, really far.

Oct 22, 2012

The picture above is funny.

But for me it is also one of those examples that make me sad about the outlook for AI and for Computer Vision. What would it take for a computer to understand this image as you or I do? I challenge you to think explicitly of all the pieces of knowledge that have to fall in place for it to make sense. Here is my short attempt:

- You recognize it is an image of a bunch of people and you understand they are in a hallway
- You recognize that there are 3 mirrors in the scene so some of those people are "fake" replicas from different viewpoints.
- You recognize Obama from the few pixels that make up his face. It helps that he is in his suit and that he is surrounded by other people with suits.
- You recognize that there's a person standing on a scale, even though the scale occupies only very few white pixels that blend with the background. But, you've used the person's pose and knowledge of how people interact with objects to figure it out.
- You recognize that Obama has his foot positioned just slightly on top of the scale. Notice the language I'm using: It is in terms of the 3D structure of the scene, not the position of the leg in the 2D coordinate system of the image.
- You know how physics works: Obama is leaning in on the scale, which applies a force on it. Scale measures force that is applied on it, that's how it works => it will over-estimate the weight of the person standing on it.
- The person measuring his weight is not aware of Obama doing this. You derive this because you know his pose, you understand that the field of view of a person is finite, and you understand that he is not very likely to sense the slight push of Obama's foot.
- You understand that people are self-conscious about their weight. You also understand that he is reading off the scale measurement, and that shortly the over-estimated weight will confuse him because it will probably be much higher than what he expects. In other words, you reason about implications of the events that are about to unfold seconds after this photo was taken, and especially about the thoughts and how they will develop inside people's heads. You also reason about what pieces of information are available to people.
- There are people in the back who find the person's imminent confusion funny. In other words you are reasoning about state of mind of people, and their view of the state of mind of another person. That's getting frighteningly meta.
- Finally, the fact that the perpetrator here is the president makes it maybe even a little more funnier. You understand what actions are more or less likely to be undertaken by different people based on their status and identity.

The state of Computer Vision and AI: we are really, really far.

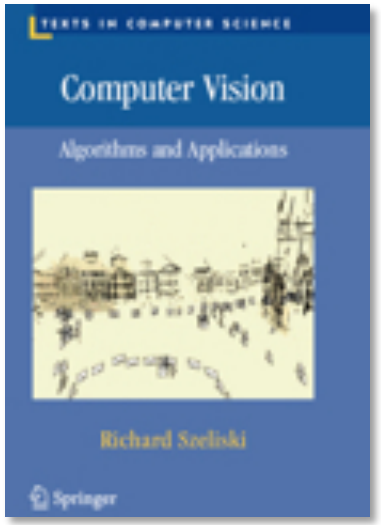
Oct 22, 2012

The picture above is funny.

But for me it is also one of those examples that make me sad about the outlook for AI and for Computer Vision. What would it take for a computer to understand this image as you or I do? I challenge you to think explicitly of all the pieces of knowledge that have to fall in place for it to make sense. Here is my short attempt:

- You recognize it is an image of a bunch of people and you understand they are in a hallway
- You recognize that there are 3 mirrors in the scene so some of those people are "fake" replicas from different viewpoints.
- You recognize Obama from the few pixels that make up his face. It helps that he is in his suit and that he is surrounded by other people with suits.
- You recognize that there's a person standing on a scale, even though the scale occupies only very few white pixels that blend with the background. But, you've used the person's pose and knowledge of how people interact with objects to figure it out.
- You recognize that Obama has his foot positioned just slightly on top of the scale. Notice the language I'm using: It is in terms of the 3D structure of the scene, not the position of the leg in the 2D coordinate system of the image.
- You know how physics works: Obama is leaning in on the scale, which applies a force on it. Scale measures force that is applied on it, that's how it works => it will over-estimate the weight of the person standing on it.
- The person measuring his weight is not aware of Obama doing this. You derive this because you know his pose, you understand that the field of view of a person is finite, and you understand that he is not very likely to sense the slight push of Obama's foot.
- You understand that people are self-conscious about their weight. You also understand that he is reading off the scale measurement, and that shortly the over-estimated weight will confuse him because it will probably be much higher than what he expects. In other words, you reason about implications of the events that are about to unfold seconds after this photo was taken, and especially about the thoughts and how they will develop inside people's heads. You also reason about what pieces of information are available to people.
- There are people in the back who find the person's imminent confusion funny. In other words you are reasoning about state of mind of people, and their view of the state of mind of another person. That's getting frighteningly meta.
- Finally, the fact that the perpetrator here is the president makes it maybe even a little more funnier. You understand what actions are more or less likely to be undertaken by different people based on their status and identity.

Course Logistics



Course webpage / Syllabus:

https://facultyweb.cs.wwu.edu/~wehrwes/courses/csci497_19w/

(link is also available on Syllabus page of Canvas)

Textbook:

Rick Szeliski, *Computer Vision: Algorithms and Applications*

online at: http://szeliski.org/Book/

Course Logistics

Announcements/grades via Canvas

Q&A via Piazza – you'll get an email invite to join

Assessment (tentative)

- 5(?) programming projects, all in Python
  - Distributed and submitted via Github
  - Some will be done in pairs, some individually
- Midterm and final exams
- Possible written homeworks
- Possible quick quizzes at the beginning of class