CS 4476 / 6476: Computer Vision

Instructor: James Hays

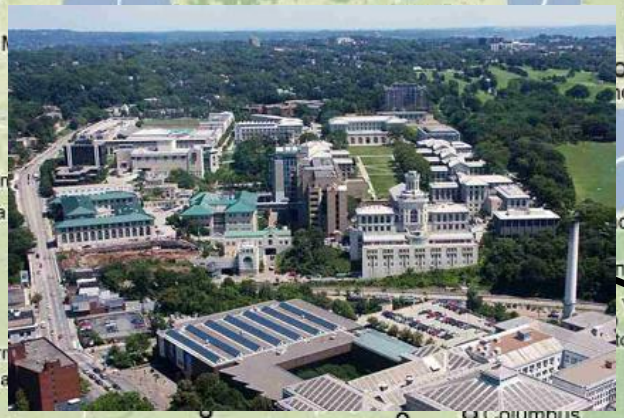
Tas: Shray Bansal, Huda Alamri, Zhaoyang Lv, Varun Agrawal, Cusuh Ham

Image by kirkh.deviantart.com

Today's Class

- Course enrollment
- Who am I?
- What is Computer Vision?
- Specifics of this course
- Questions

A bit about me



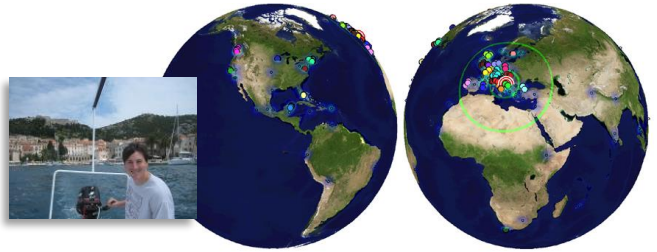
My recent research

Internet Data

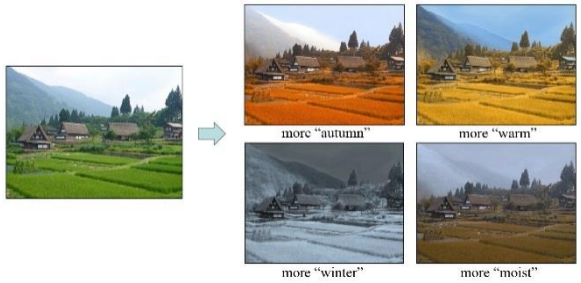
Crowd Data

Image Restoration

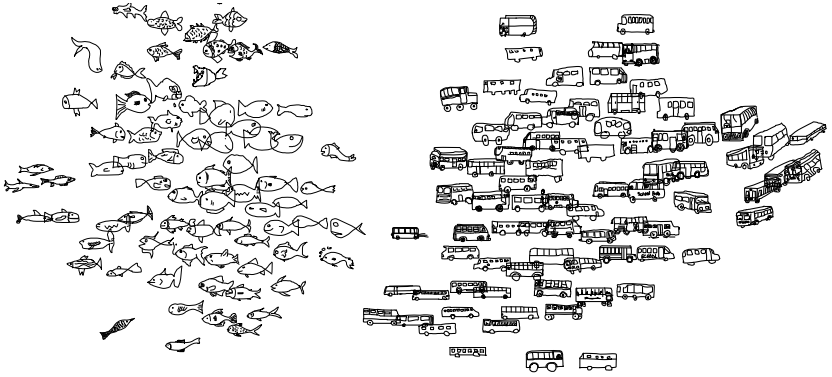
Image Geolocalization



High-level Image Editing



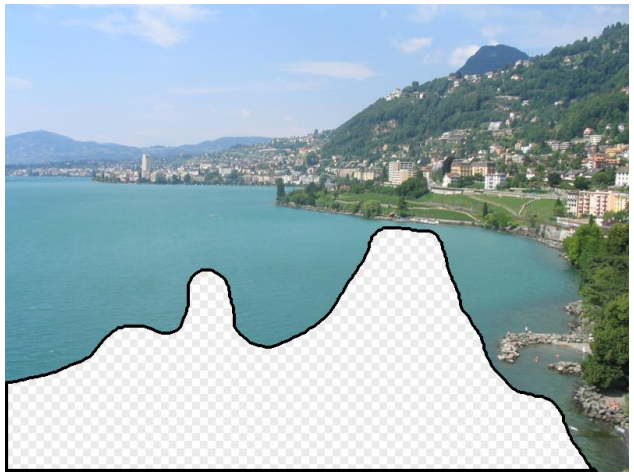
How do Humans Sketch Objects?

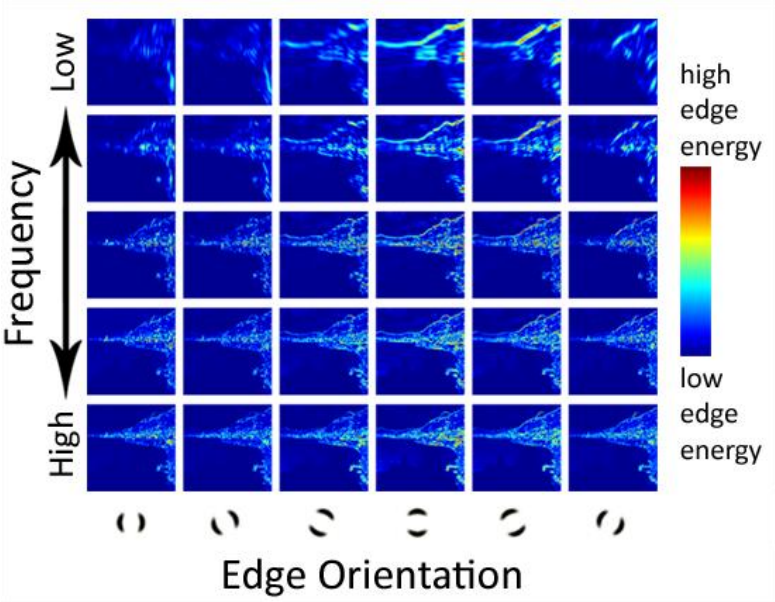


Input Image

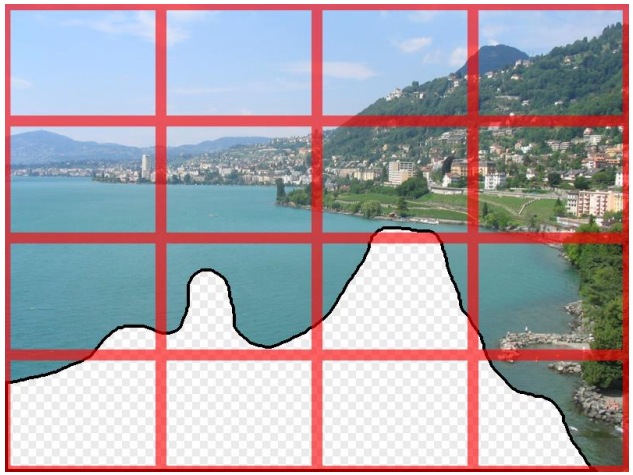
Hays and Efros. **Scene Completion Using Millions of Photographs**. SIGGRAPH 2007 and Communications of ACM October 2008.

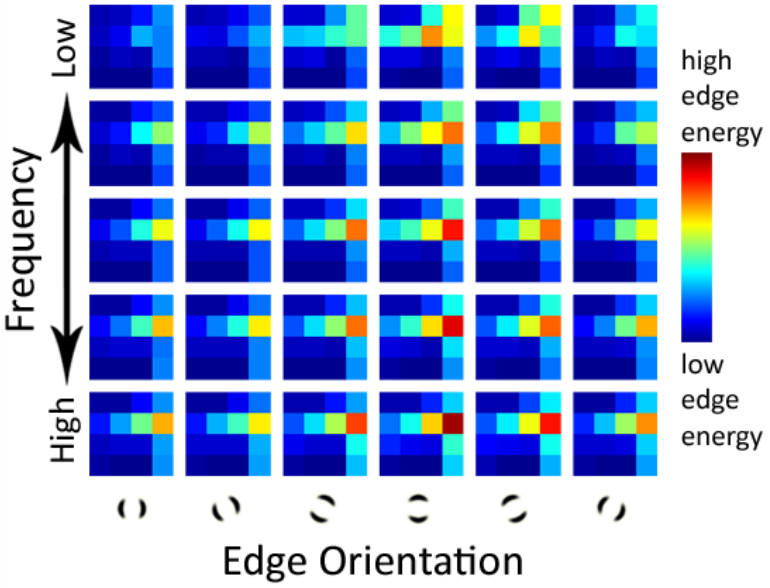
Scene Descriptor





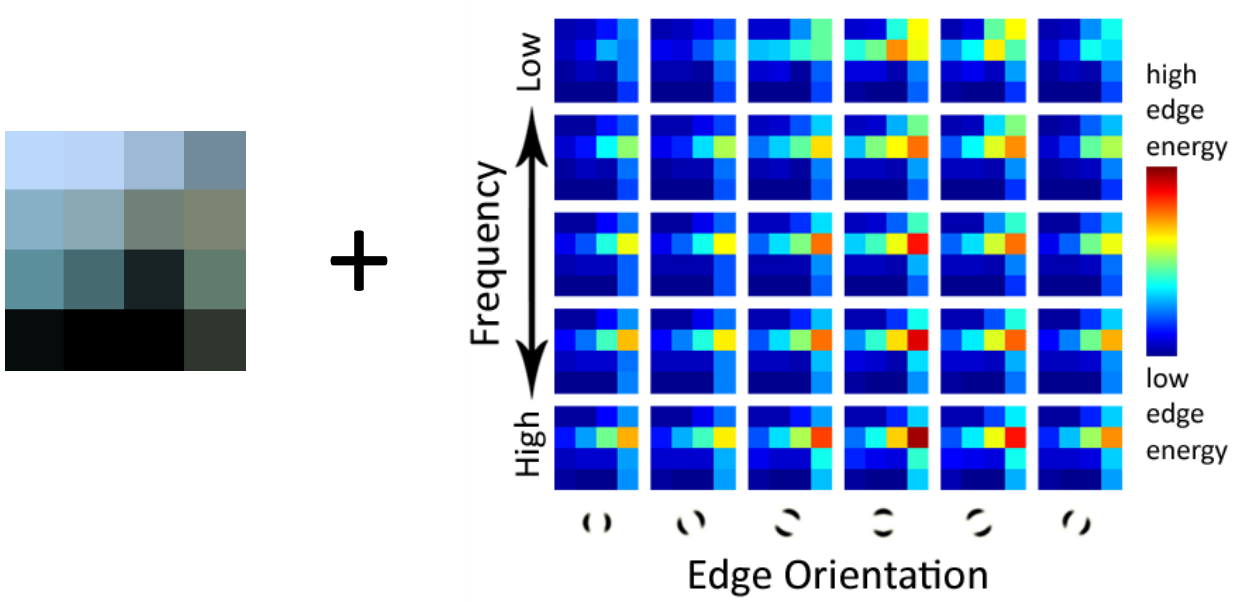
Scene Descriptor



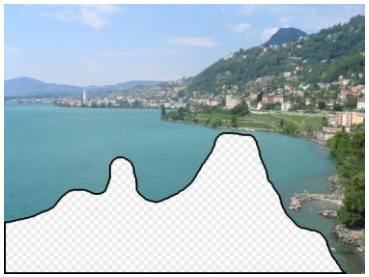


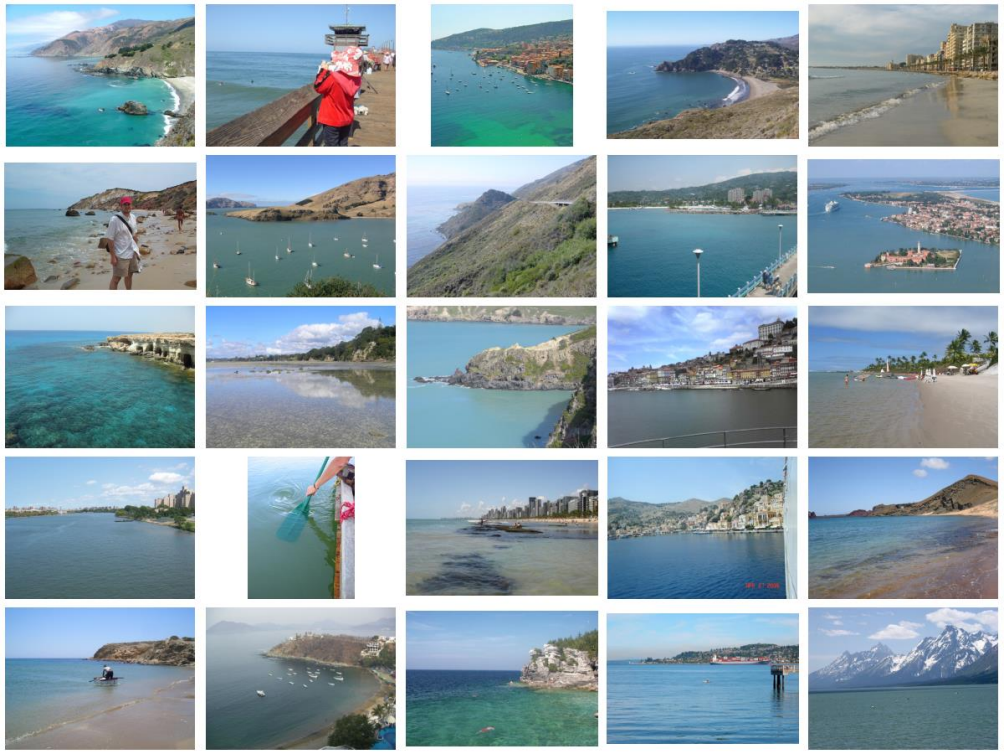
Scene Gist Descriptor
(Oliva and Torralba 2001)

Scene Descriptor



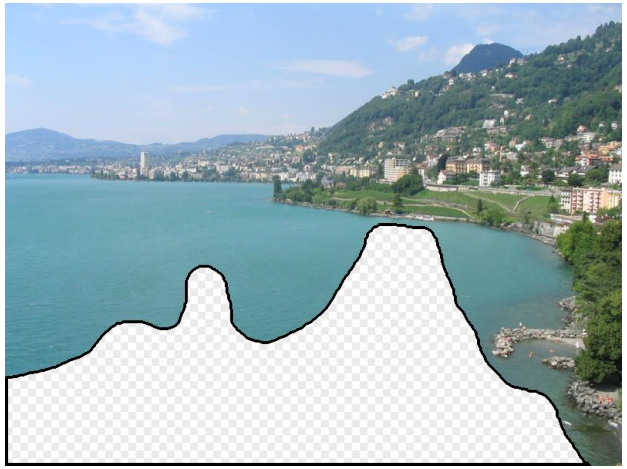
Scene Gist Descriptor
(Oliva and Torralba 2001)





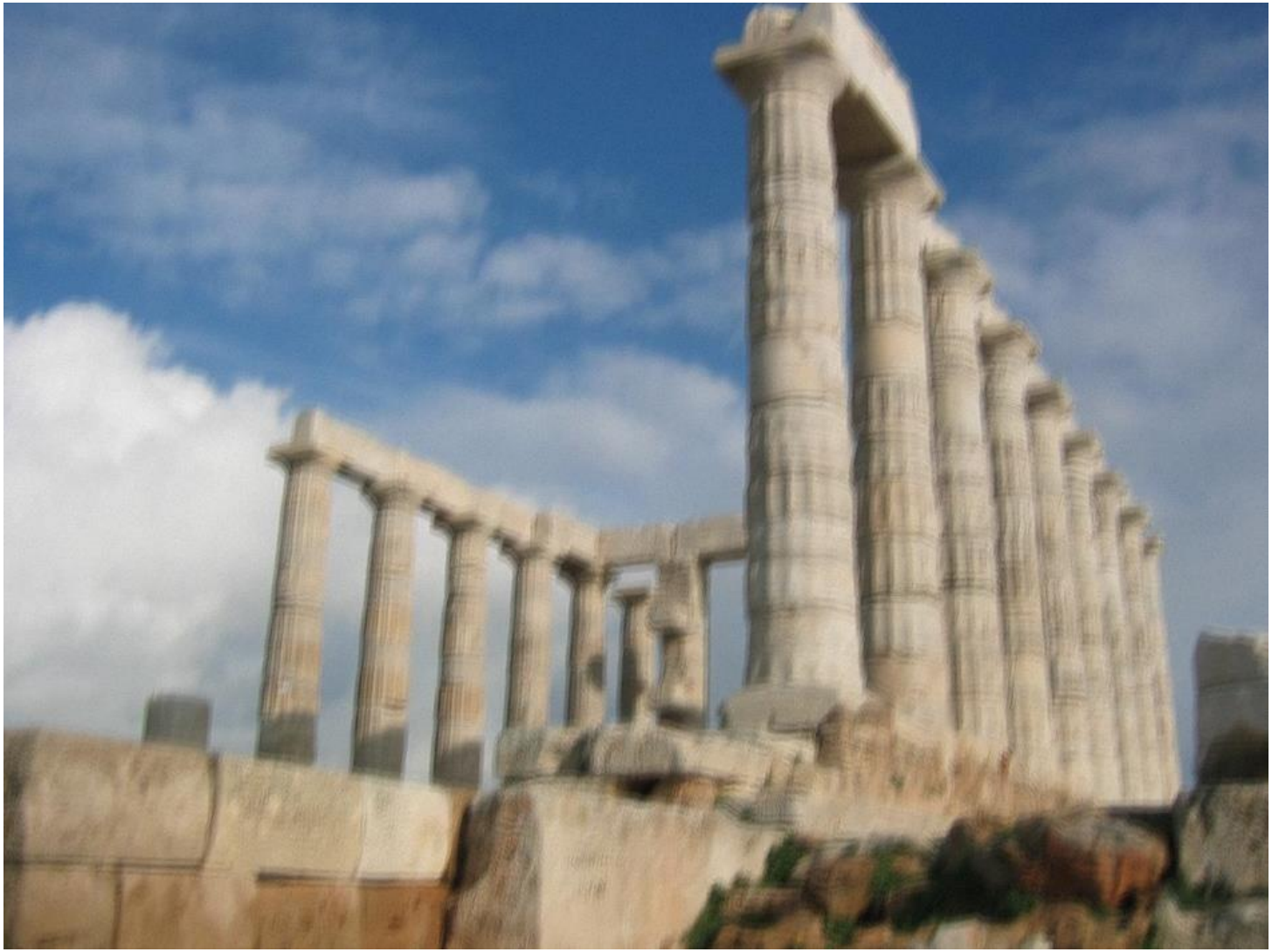
… 200 total

Context Matching




Graph cut + Poisson blending

Repairing Camera Shake



Our-Synth + Zoran and Weiss


Original

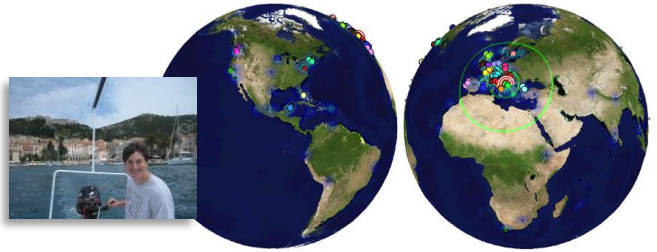
My recent research

Internet Data

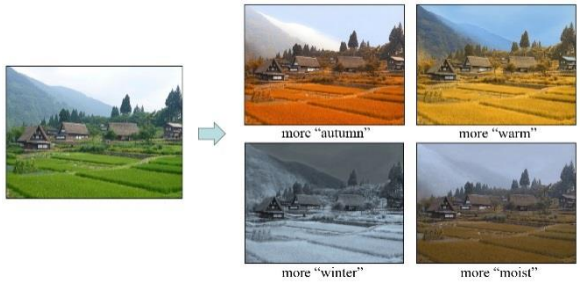
Crowd Data

Image Restoration

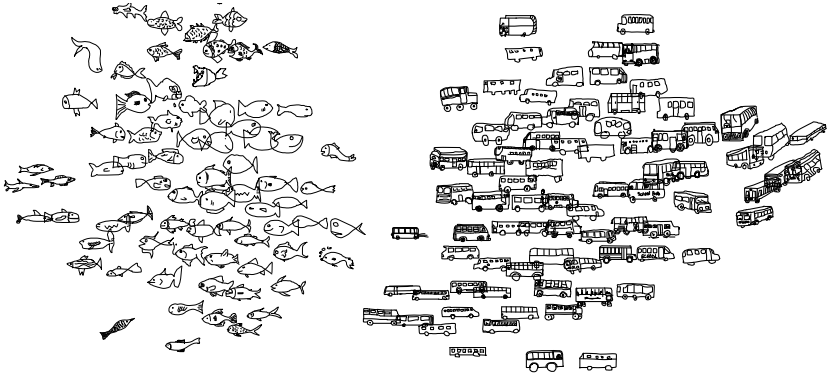
Image Geolocalization



High-level Image Editing

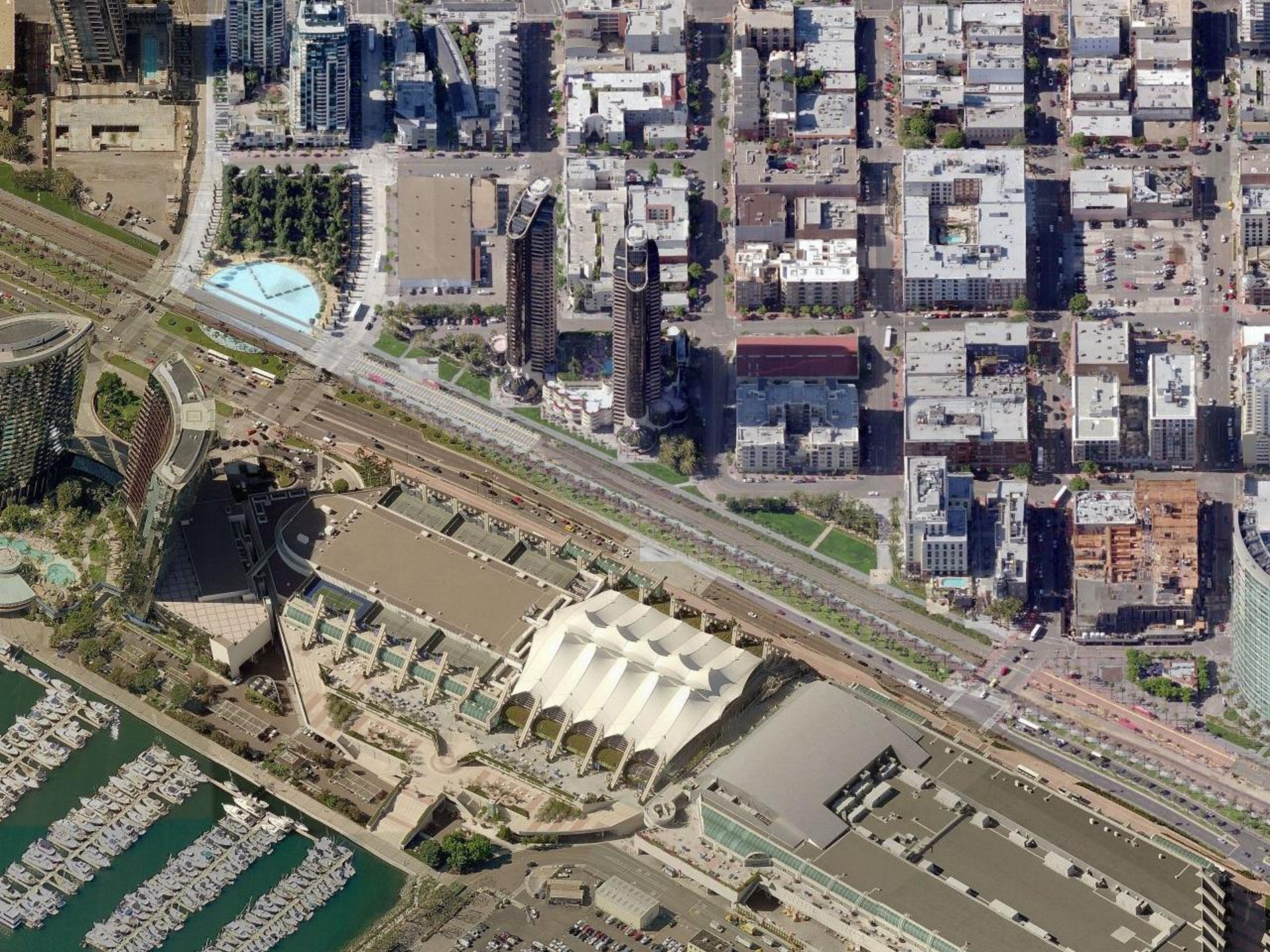


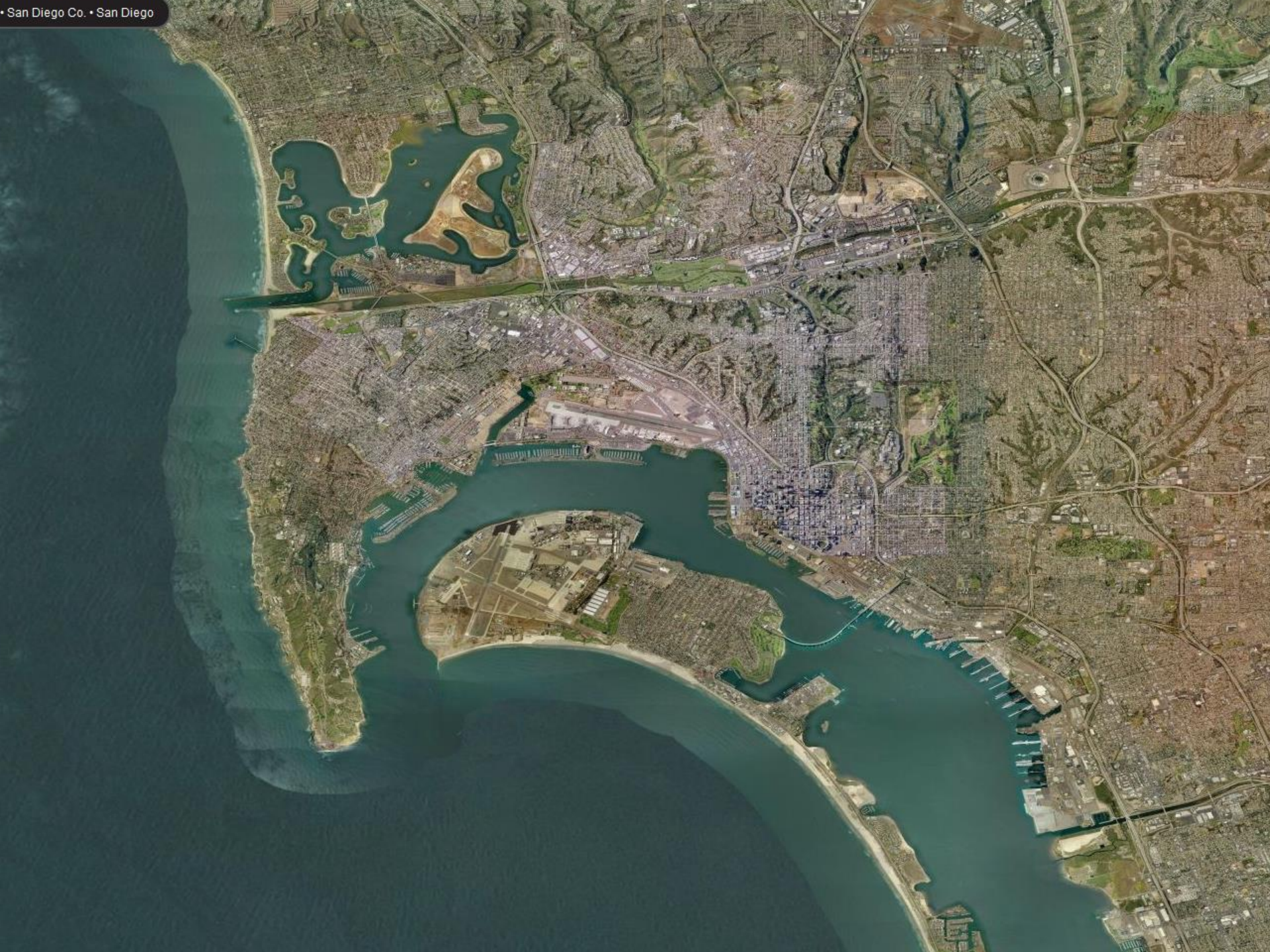
How do Humans Sketch Objects?



Learning Deep Representations for Ground-to-Aerial Geolocalization.

Lin, Cui, Belongie, and Hays. **Learning Deep Representations for Ground-to-Aerial Geolocalization**. CVPR 2015 (Oral).

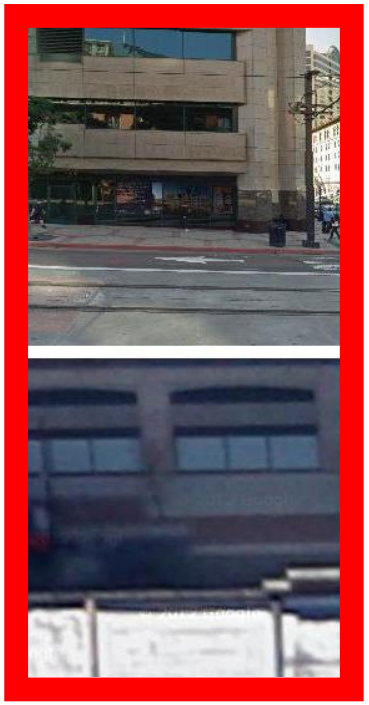


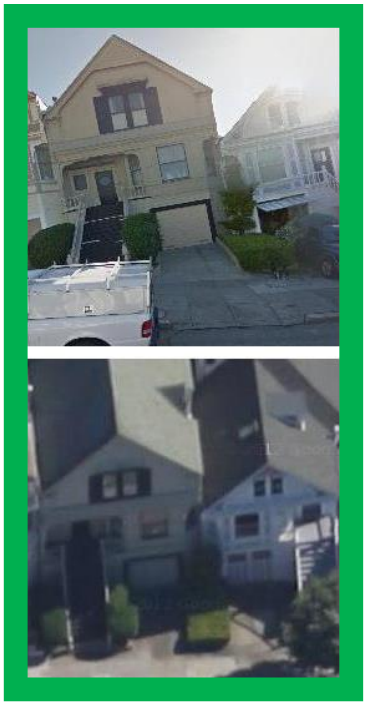

• San Diego Co. • San Diego

Are these the same location?

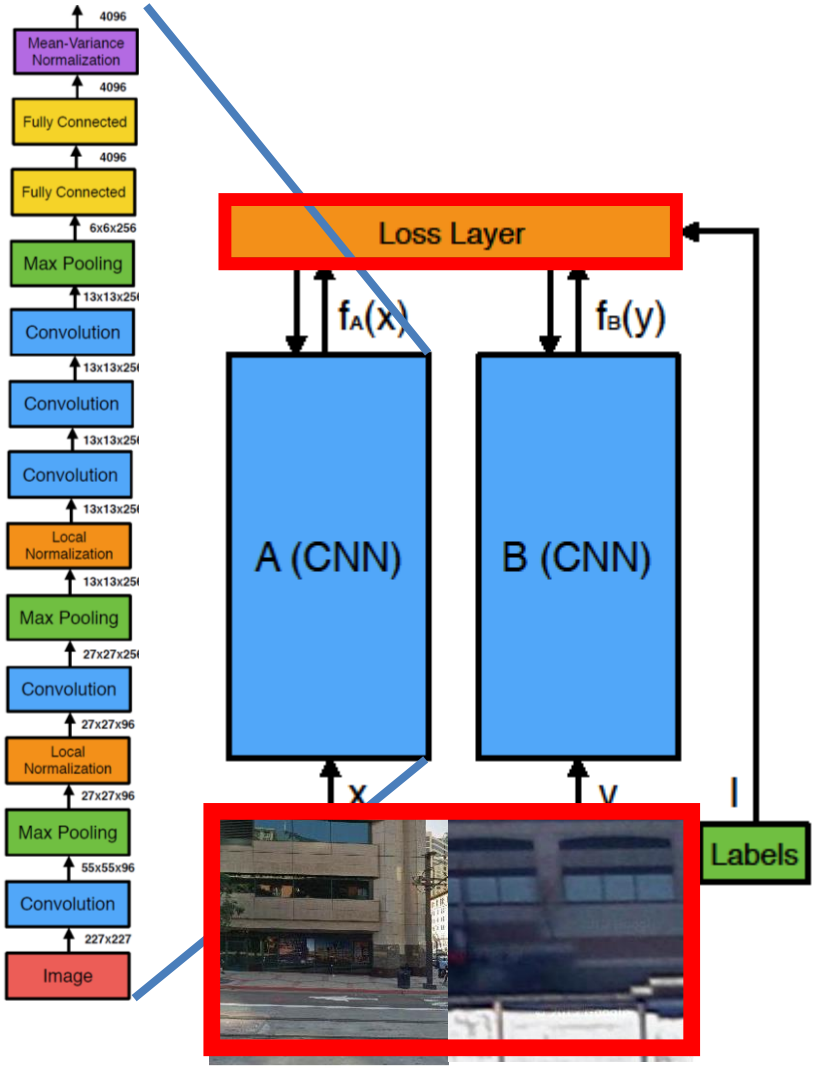
Ground

Aerial

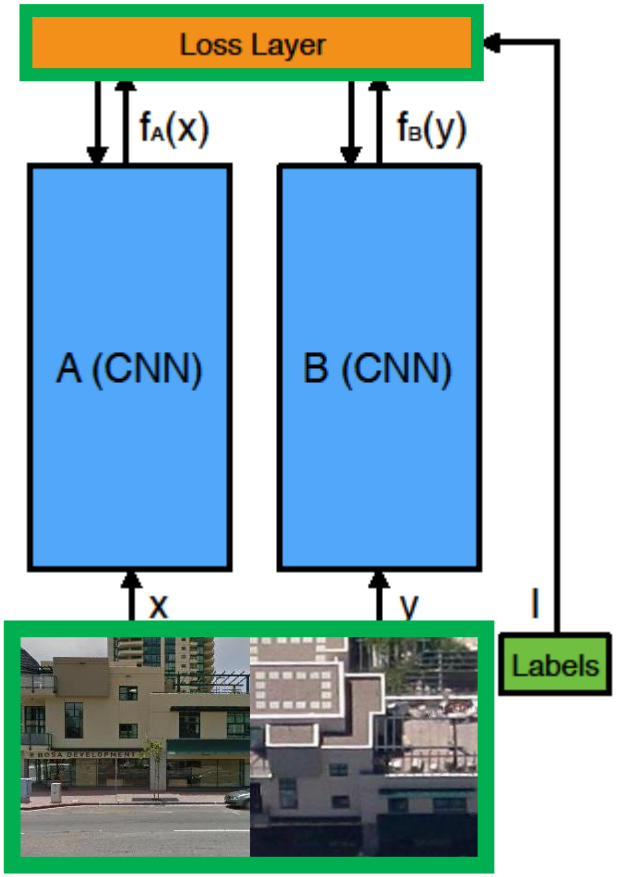


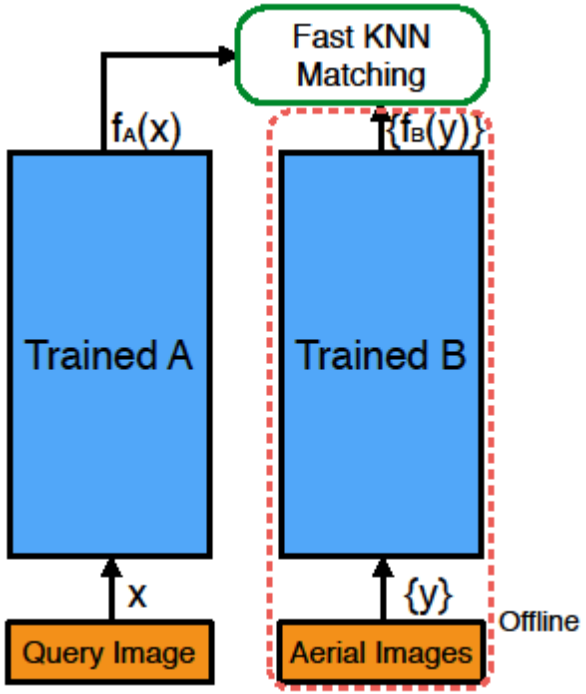


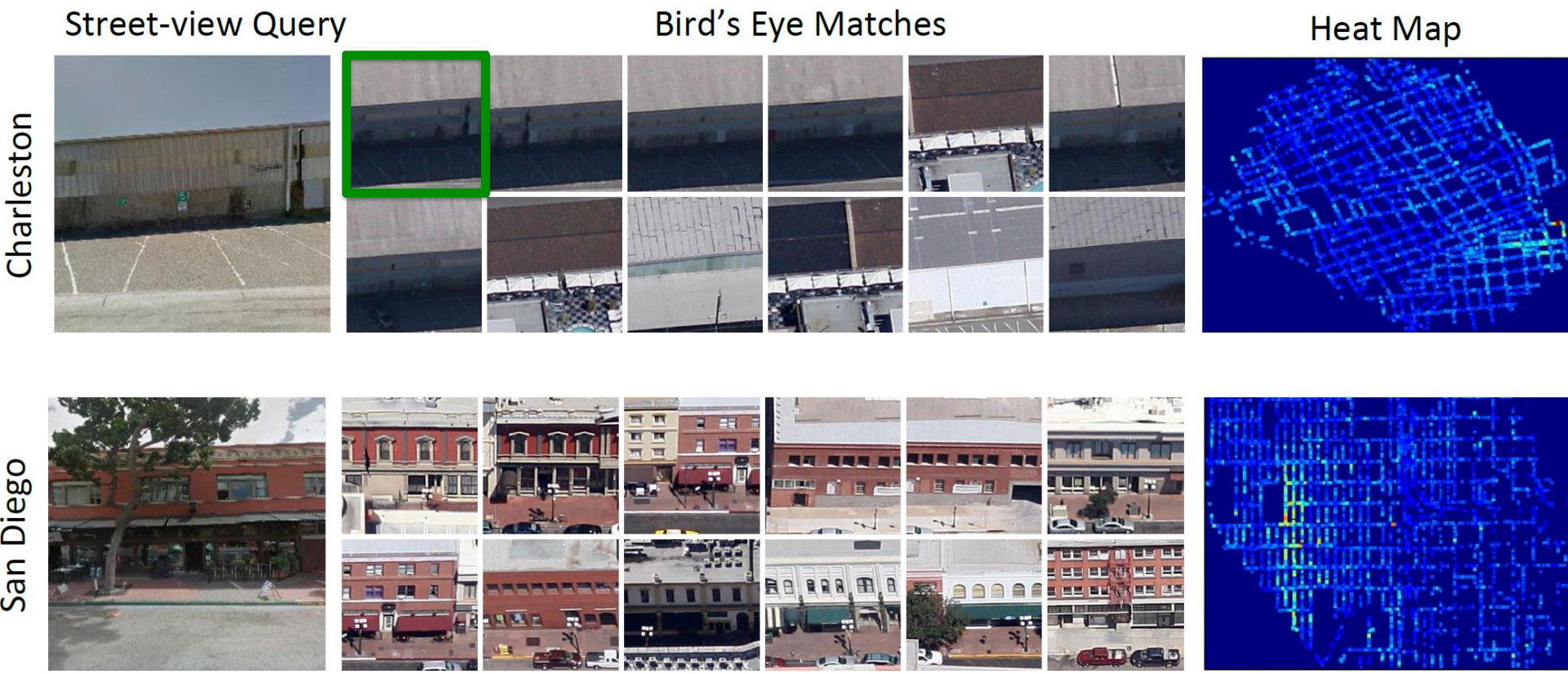
“Siamese” ConvNet for Ground-to-Aerial Matching



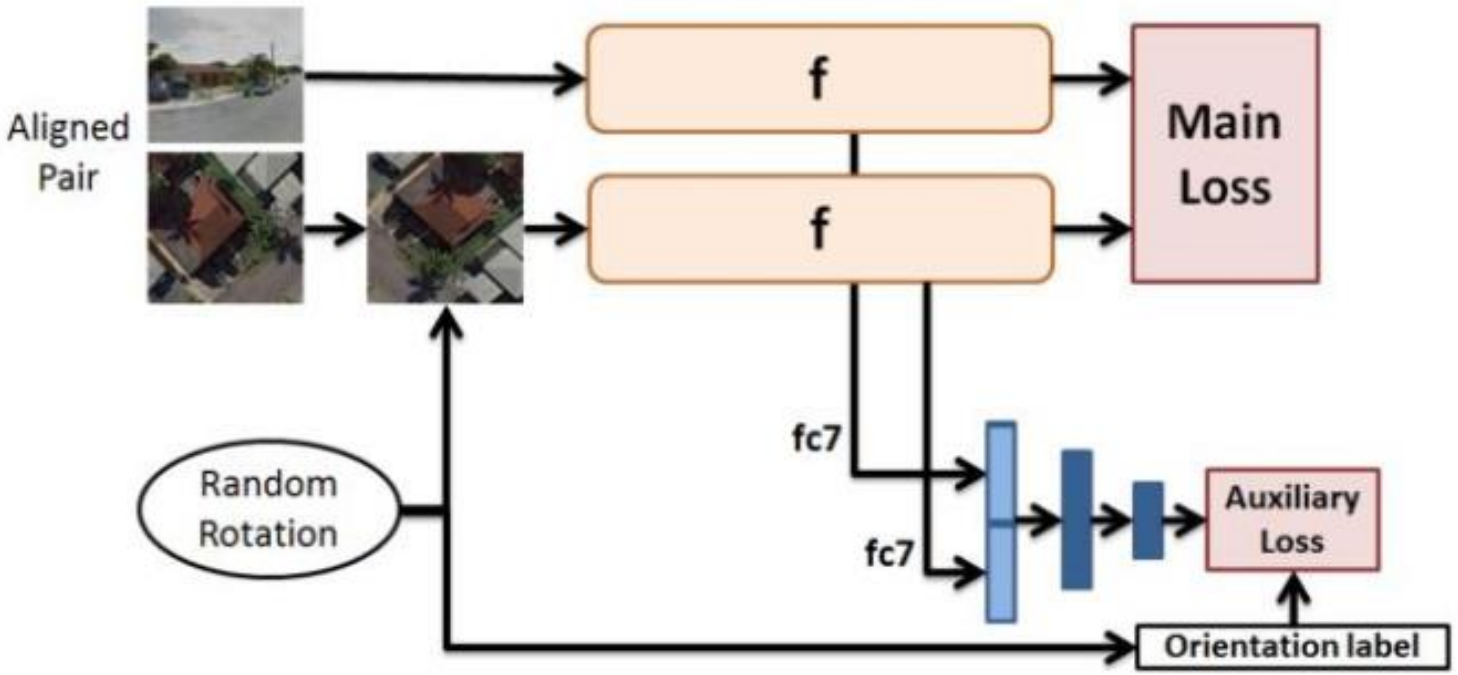
“Siamese” ConvNet for Ground-to-Aerial Matching

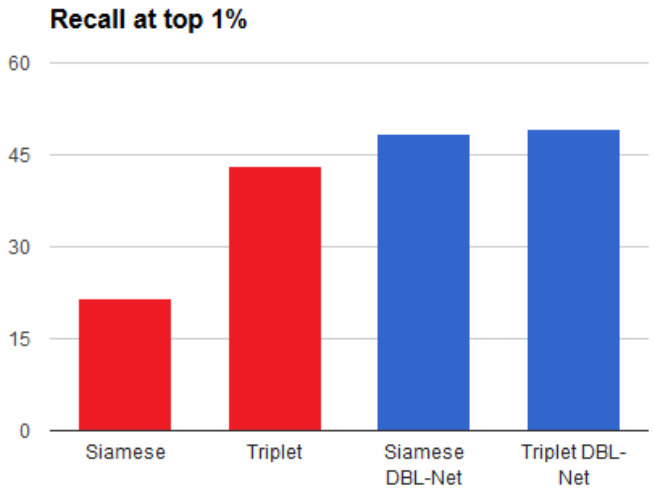





Street-view Query
Bird's Eye Matches
Heat Map
Charleston
San Diego

ECCV 2016 Extensions by Nam Vo





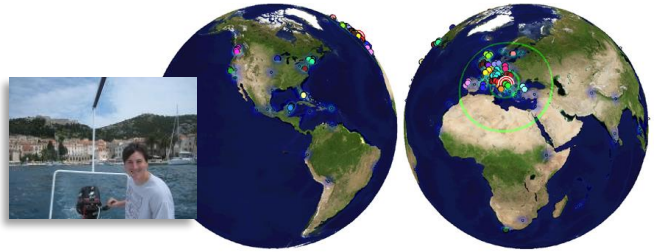
My recent research

Internet Data

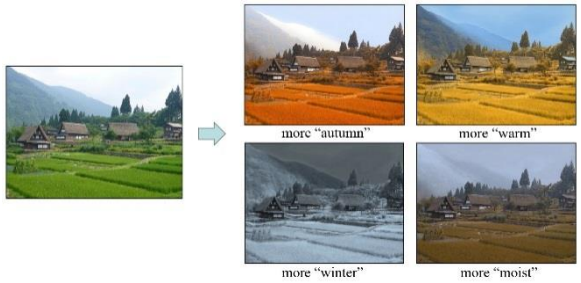
Crowd Data

Image Restoration

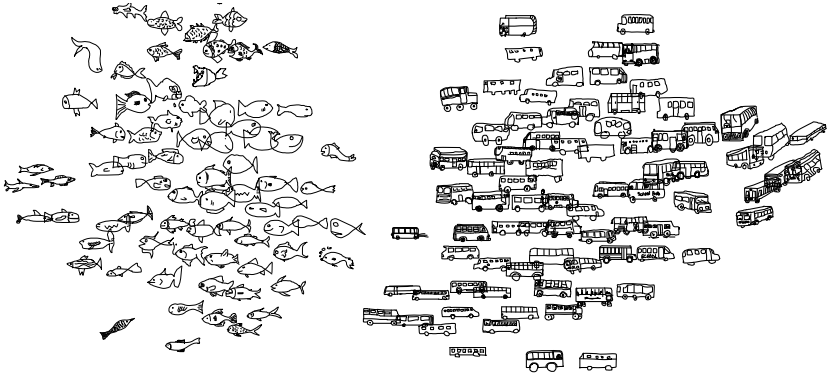
Image Geolocalization



High-level Image Editing



How do Humans Sketch Objects?



Input image (courtesy minque)

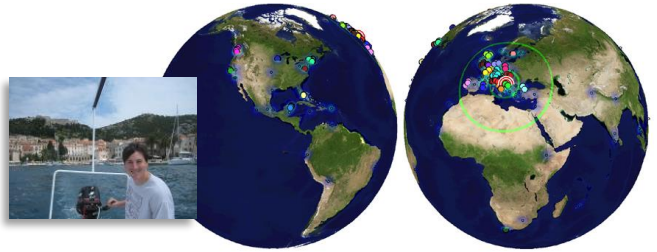
My recent research

Internet Data

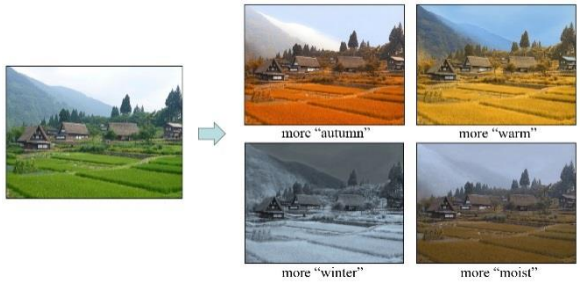
Crowd Data

Image Restoration

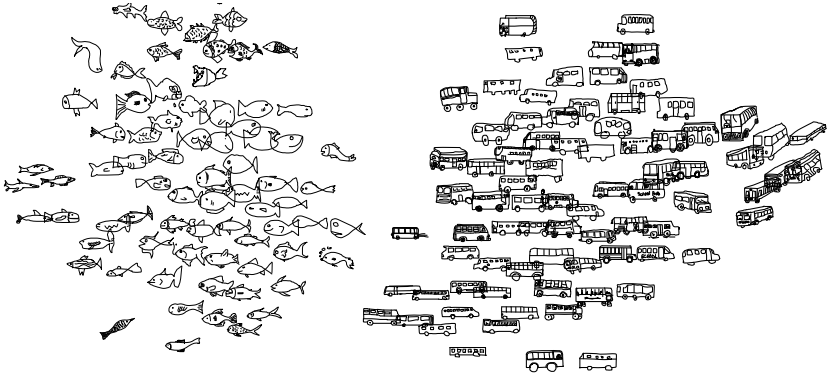
Image Geolocalization



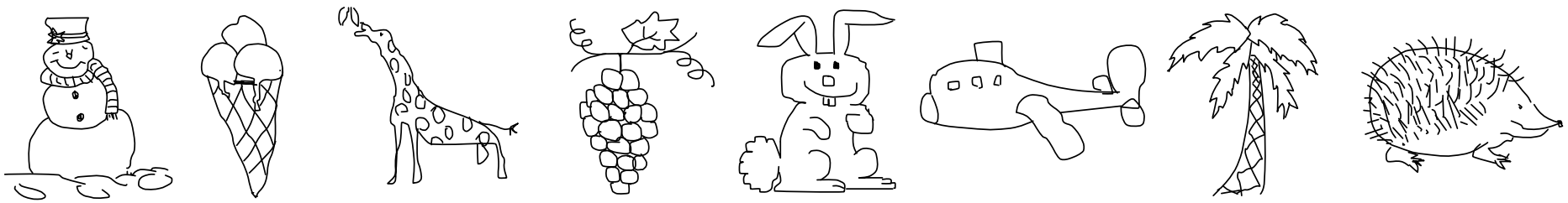
High-level Image Editing



How do Humans Sketch Objects?



How Do Humans Sketch Objects?



Mathias Eitz, James Hays, and Marc Alexa.
How Do Humans Sketch Objects? Siggraph 2012

Sketches Are Important

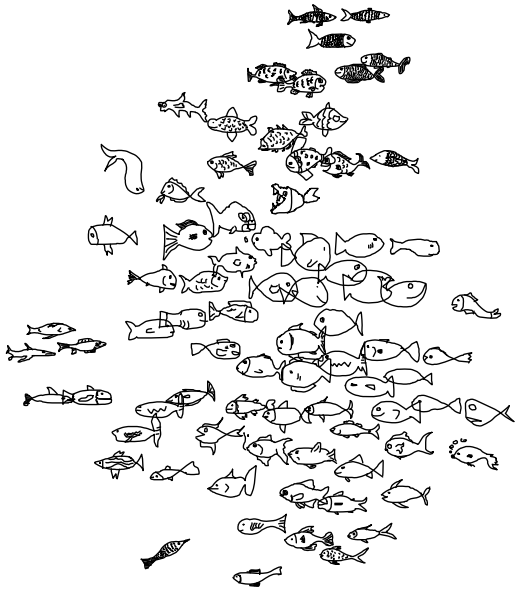
17,000 years ago (Lascaux, France)

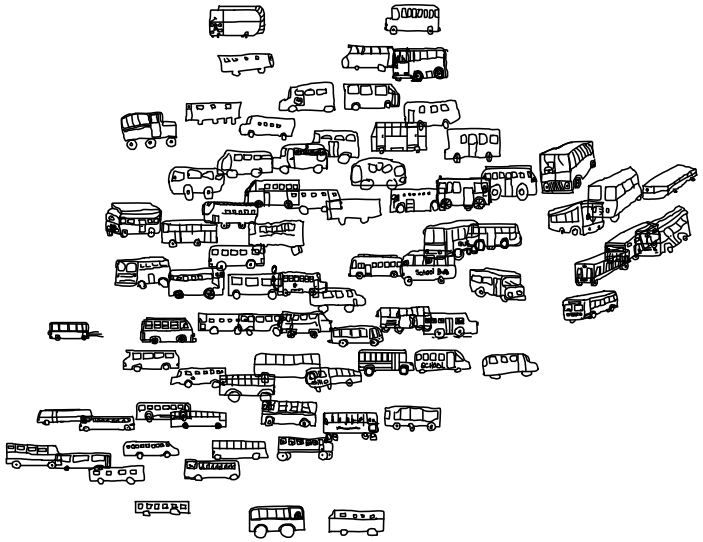
Sketches Are Important

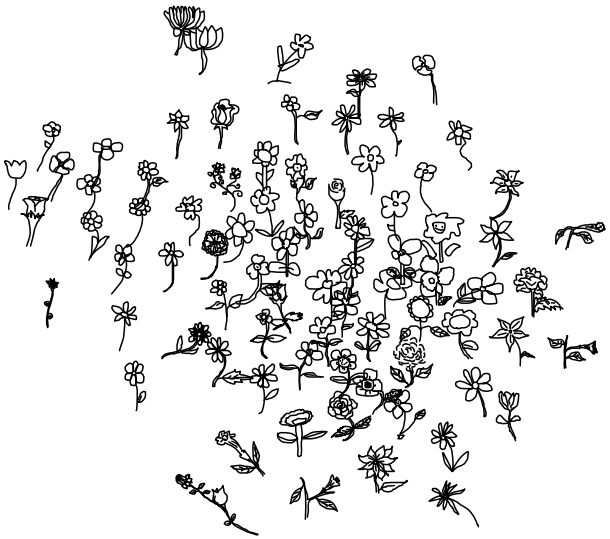
Today

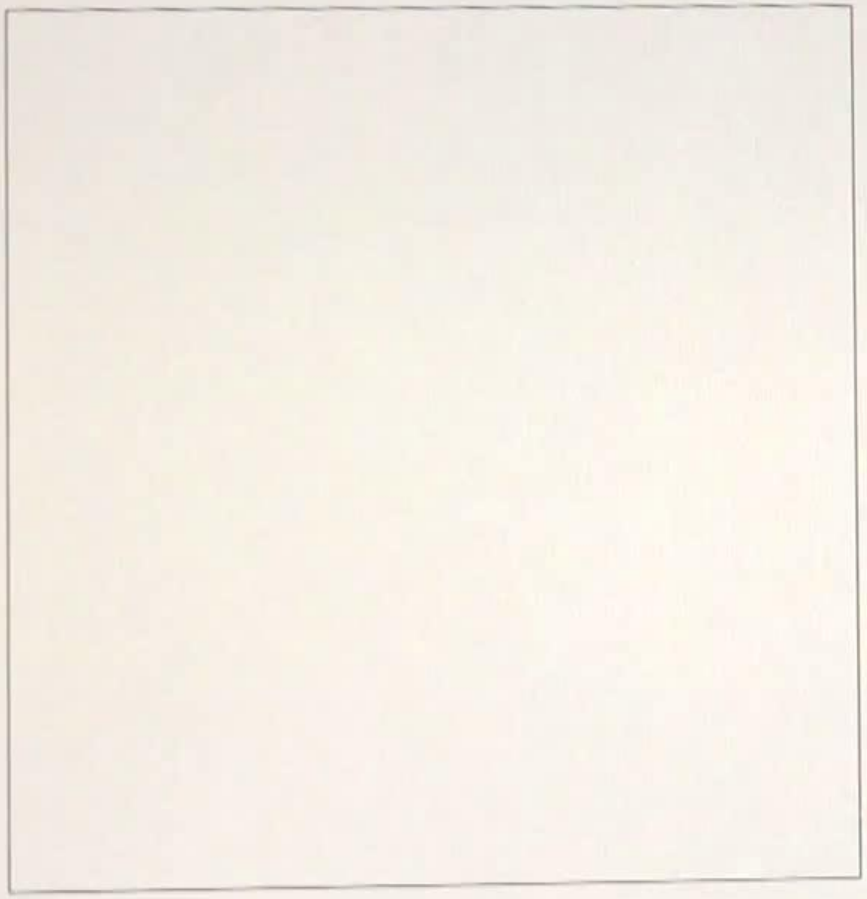
How Do Humans Sketch Objects?

- 20,000 sketches in 250 categories
  - 1,350 unique participants, 741 hours drawing time









Siggraph 2016

http://sketchy.eye.gatech.edu/

What is Computer Vision?

- What are examples of computer vision being used in the world?

Computer Vision

Make computers understand images and video.

What kind of scene?

Where are the cars?

How far is the building?

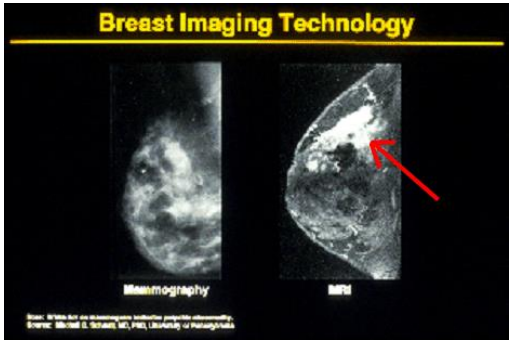
…

Vision is really hard

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

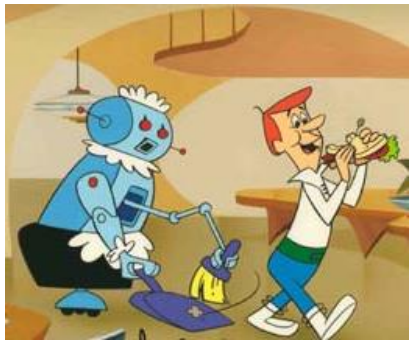
Why computer vision matters

Safety



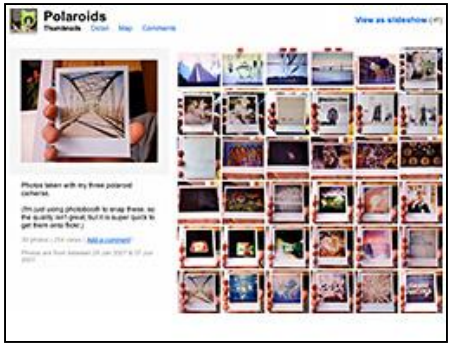
Health

Security



Comfort

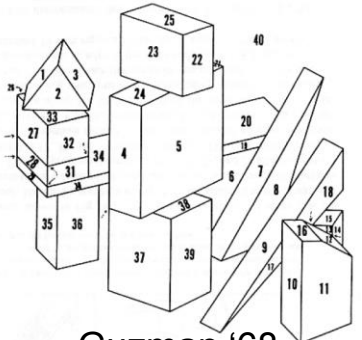
Fun



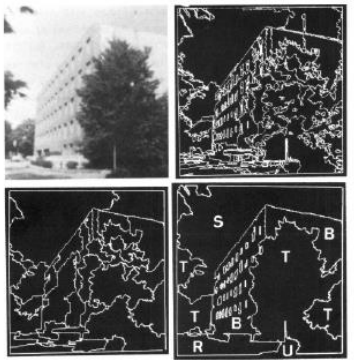
Access

Ridiculously brief history of computer vision

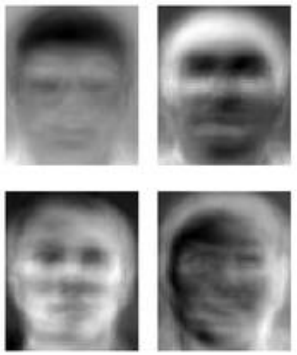
- 1966: Minsky assigns computer vision as an undergrad summer project
- 1960's: interpretation of synthetic worlds
- 1970's: some progress on interpreting selected images
- 1980's: ANNs come and go; shift toward geometry and increased mathematical rigor
- 1990's: face recognition; statistical analysis in vogue
- 2000's: broader recognition; large annotated datasets available; video processing starts
- 2010's: Deep learning with ConvNets
- 2030's: robot uprising?



Guzman '68



Ohta Kanade '78



Turk and Pentland '91

How vision is used now

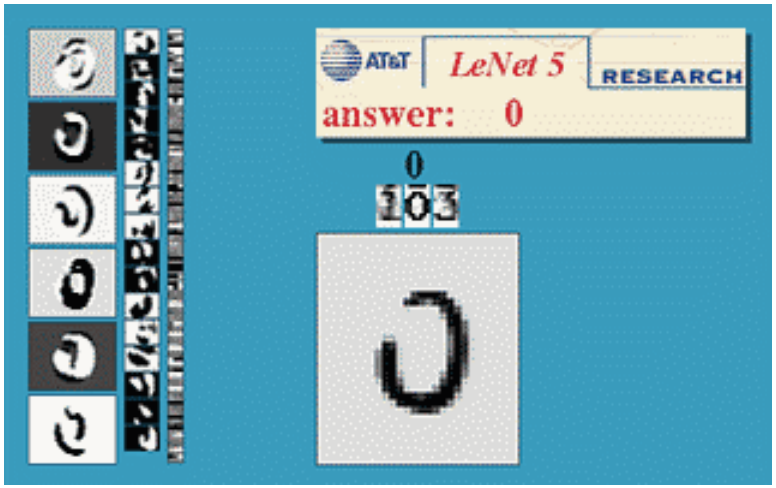
- Examples of recent real world applications

Some of the following slides by Steve Seitz

Optical character recognition (OCR)

Technology to convert scanned docs to text

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



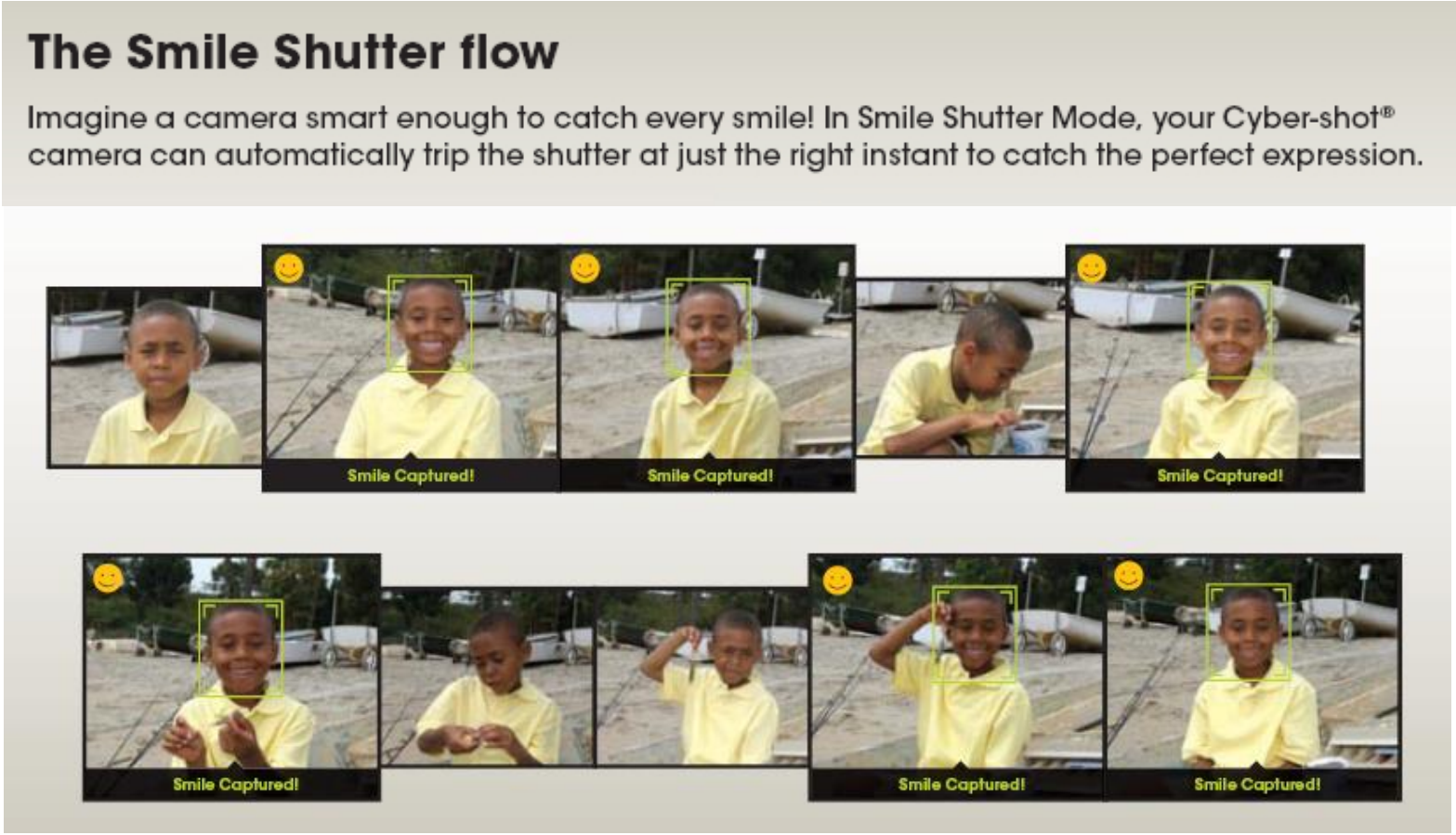
Digit recognition, AT&T labs
http://www.research.att.com/~yann/

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

Face detection

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

Smile detection



[Sony Cyber-shot® T70 Digital Still Camera]

3D from thousands of images

Building Rome in a Day: Agarwal et al. 2009

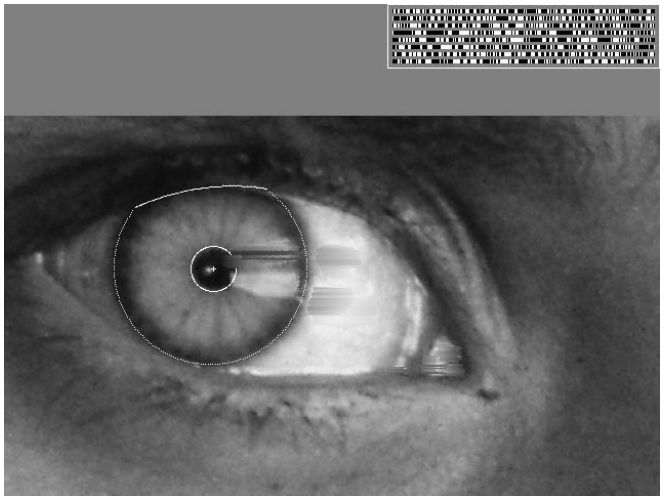
Object recognition (in supermarkets)

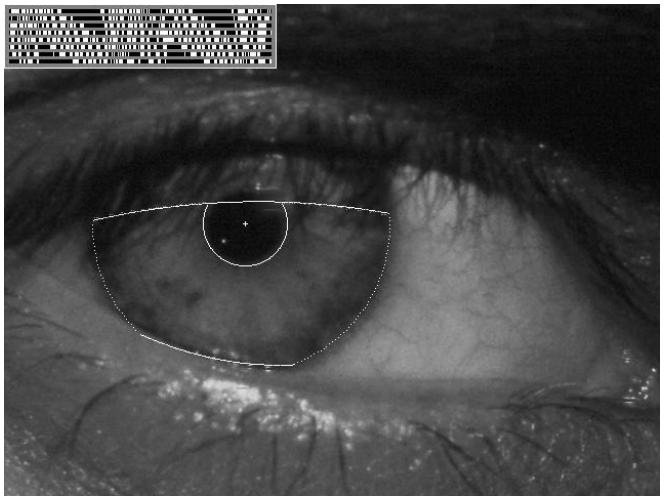
LaneHawk by EvolutionRobotics

“A smart camera is flush-mounted in the checkout lane, continuously watching for items. When an item is detected and recognized, the cashier verifies the quantity of items that were found under the basket, and continues to close the transaction. The item can remain under the basket, and with LaneHawk,you are assured to get paid for it… “

Vision-based biometrics

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





Login without a password…

Fingerprint scanners on many new laptops, other devices

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

Object recognition (in mobile phones)

[Point & Find](), [Nokia]()

[Google Goggles]()

Special effects: shape capture

The Matrix movies, ESC Entertainment, XYZRGB, NRC

Special effects: motion capture

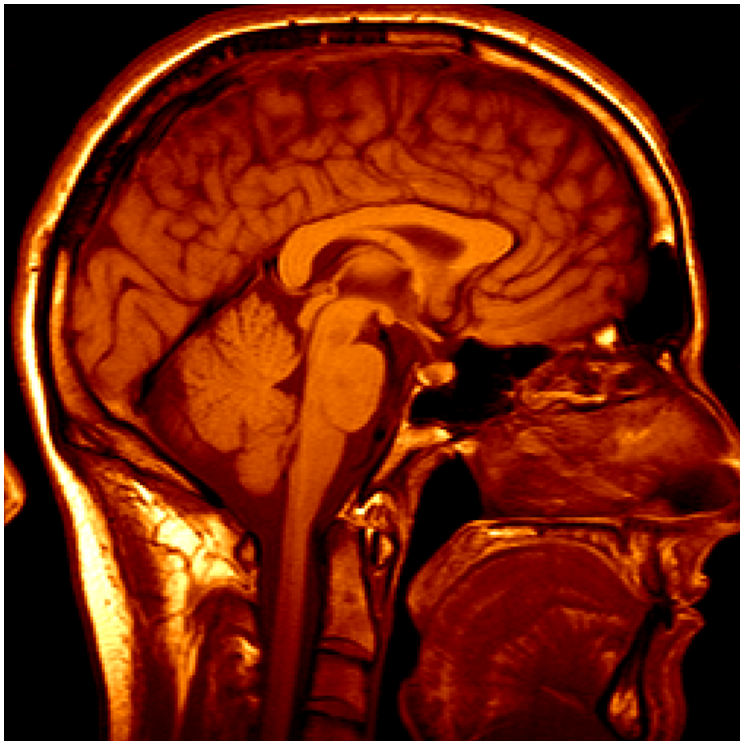
Pirates of the Carribean, Industrial Light and Magic

Sports

Sportvision first down line
Nice explanation on www.howstuffworks.com

http://www.sportvision.com/video.html

Medical imaging



3D imaging
MRI, CT

Image guided surgery
Grimson et al., MIT

Smart cars

Slide content courtesy of Amnon Shashua

- Mobileye
  - Market Capitalization: 11 Billion dollars

Google cars

Oct 9, 2010. "Google Cars Drive Themselves, in Traffic". *The New York Times*. John Markoff

June 24, 2011. "Nevada state law paves the way for driverless cars". *Financial Post*. Christine Dobby

Aug 9, 2011, "Human error blamed after Google's driverless car sparks five-vehicle crash". *The Star* (Toronto)

Ford acquires SAIPS for self-driving machine learning and computer vision tech

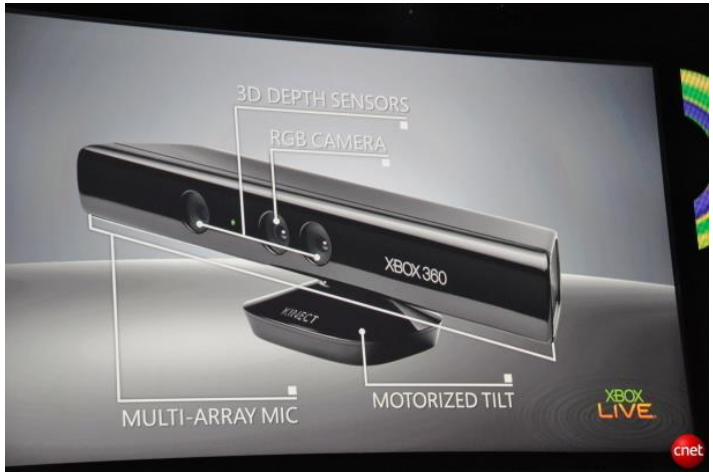
Posted Aug 16, 2016 by **Darrell Etherington** *(@etherington)*

Ford outlined a few of the ways it's aiming to ship driverless cars by 2021, and part of the plan involves acquisitions. CEO Mark Fields revealed at a press event in Palo Alto today that the automaker acquired SAIPS, an Israeli company focusing on machine learning and computer vision. It's also partnering exclusively with Nirenberg Neuroscience, to bring more "humanlike intelligence" to machine learning components of driverless car systems.

SAIPS' technology brings image and video processing algorithms, as well as deep learning tech focused on processing and classifying input signals, all key ingredients in the special sauce that makes up autonomous vehicle tech. This company's expertise should help with on-board interpretation of data captured by sensors on Ford's self-driving cars, and turning that data into usable info for the car's virtual driver system. SAIPS' offerings include detection of anomalies, persistent tracking of objects detected by sensors, and much more. The company's past clients include HP and Trax, but its partner group doesn't appear to have included much in the way of driving-specific applications.

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



Industrial robots

Vision-guided robots position nut runners on wheels

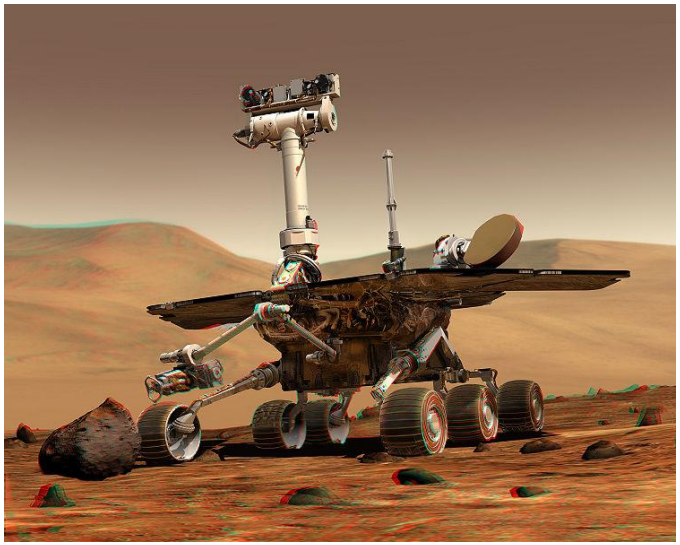
Vision in space

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

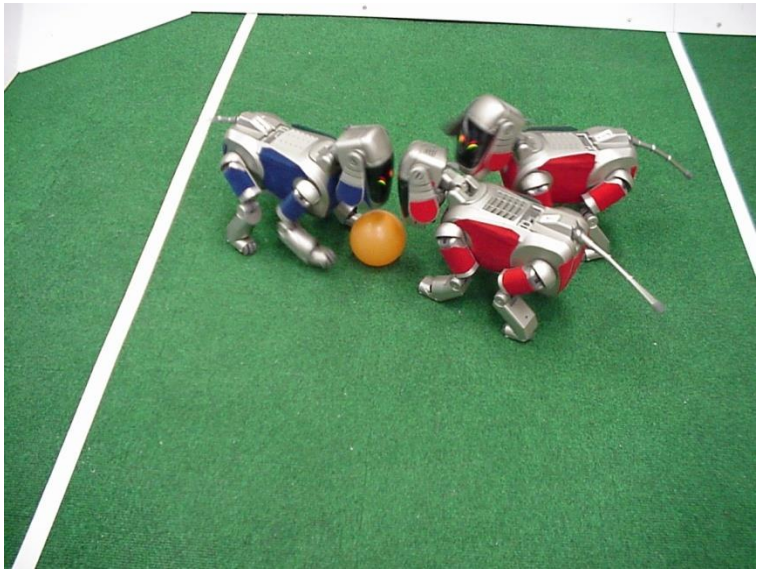
Vision systems (JPL) used for several tasks

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

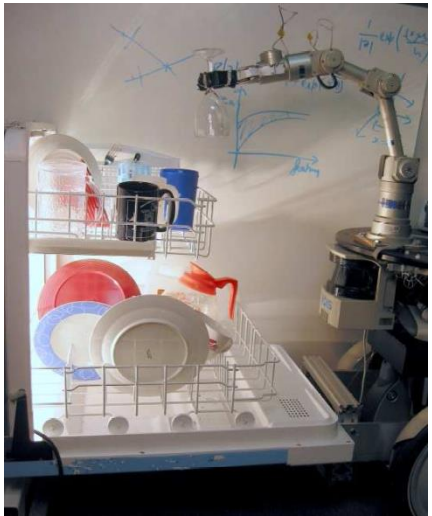
Mobile robots



NASA's Mars Spirit Rover
http://en.wikipedia.org/wiki/Spirit_rover



http://www.robocup.org/



Saxena et al. 2008
STAIR at Stanford

Amazon Prime Air

https://www.amazon.com/b?node=8037720011

Skydio

https://www.skydio.com/

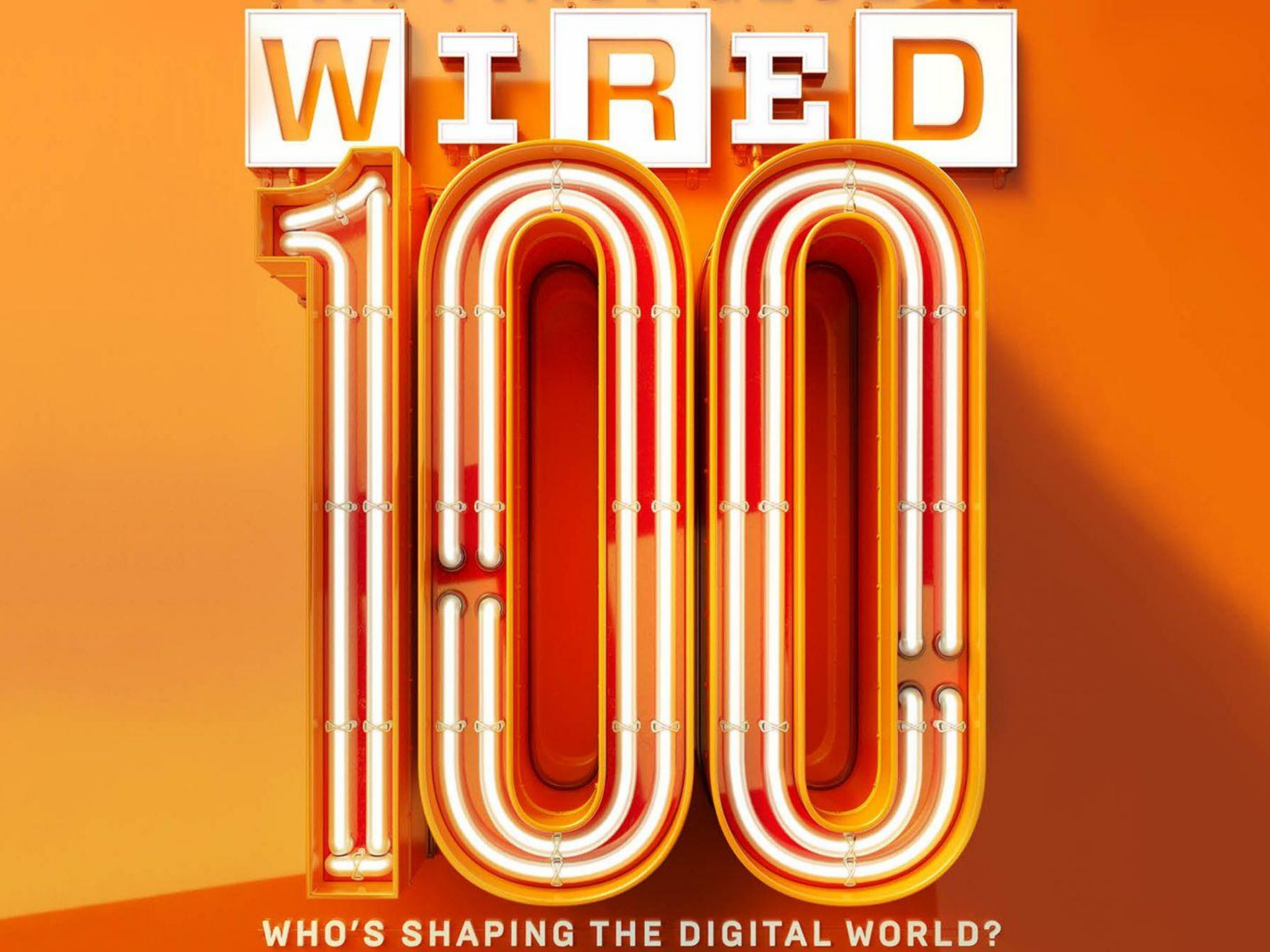
Augmented Reality and Virtual Reality

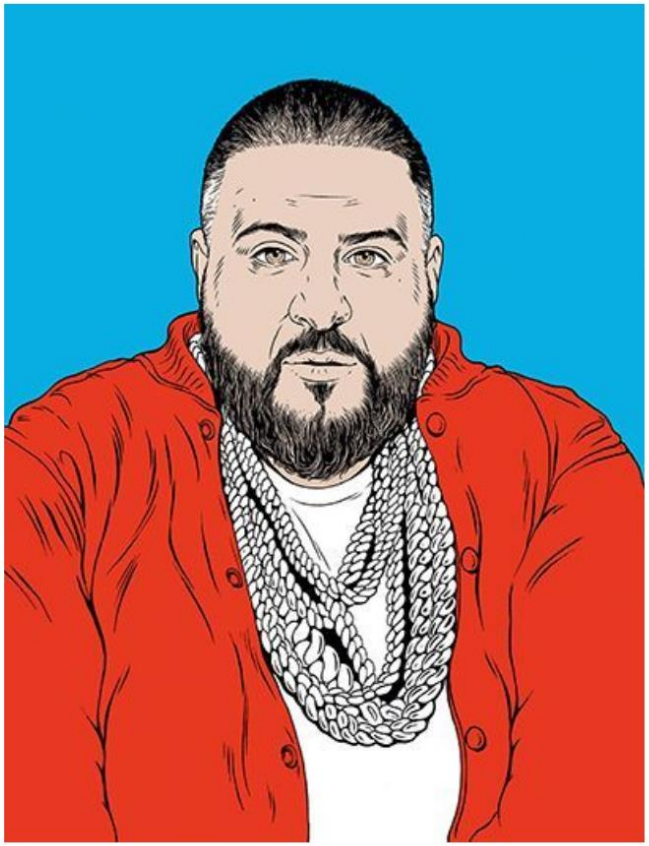
Magic Leap, Oculus, Hololens, etc.

State of the art today?

With enough training data, computer vision nearly matches human vision at most recognition tasks

Deep learning has been an enormous disruption to the field. More and more techniques are being “deepified”.


WIRED
100
WHO'S SHAPING THE DIGITAL WORLD?



DJ Khaled

Credit **Louise Zergaeng Pomeroy**

73. DJ Khaled

Snapchat icon; DJ and producer

Louisiana-born Khaled Mohamed Khaled, aka DJ Khaled, cut his musical chops in the early 00s as a host for Miami urban music radio WEDR. He proceeded to build a solid if not dazzling career as a mixtape DJ and music producer (he founded his label We The Best Music Group in 2008, and was appointed president of Def Jam South in 2009).

69. Geoffrey Hinton

Psychologist, computer scientist; researcher, Google Toronto

British-born Hinton has been dubbed the "godfather of deep learning". The Cambridge-educated cognitive psychologist and computer scientist started being an ardent believer in the potential of neural networks and deep learning in the 80s, when those technologies enjoyed little support in the wider AI community.

But he soldiered on: in 2004, with support from the Canadian Institute for Advanced Research, he launched a University of Toronto programme in neural computation and adaptive perception, where, with a group of researchers, he carried on investigating how to create computers that could behave like brains.

Hinton's work – in particular his algorithms that train multilayered neural networks – caught the attention of tech giants in Silicon Valley, which realised how deep learning could be applied to voice recognition, predictive search and machine vision.

The spike in interest prompted him to launch a free course on neural networks on e-learning platform Coursera in 2012. Today, 68-year-old Hinton is chair of machine learning at the University of Toronto and moonlights at Google, where he has been using deep learning to help build internet tools since 2013.

63. Yann Lecun

Director of AI research, Facebook, Menlo Park

LeCun is a leading expert in deep learning and heads up what, for Facebook, could be a hugely significant source of revenue: understanding its user's intentions.

62. Richard Branson

Founder, Virgin Group, London

Branson saw his personal fortune grow £550 million when Alaska Air bought Virgin America for $2.6 billion in April. He is pressing on with civilian space travel with Virgin Galactic.

61. Taylor Swift

Entertainer, Los Angeles

Credit **Google DeepMind**

Google-backed startup DeepMind Technologies has built an artificial intelligence agent that can learn to successfully play 49 classic Atari games by itself, with minimal input.

Computer Vision and Nearby Fields

- Computer Graphics: Models to Images
- Comp. Photography: Images to Images
- Computer Vision: Images to Models

Derogatory summary of computer vision:

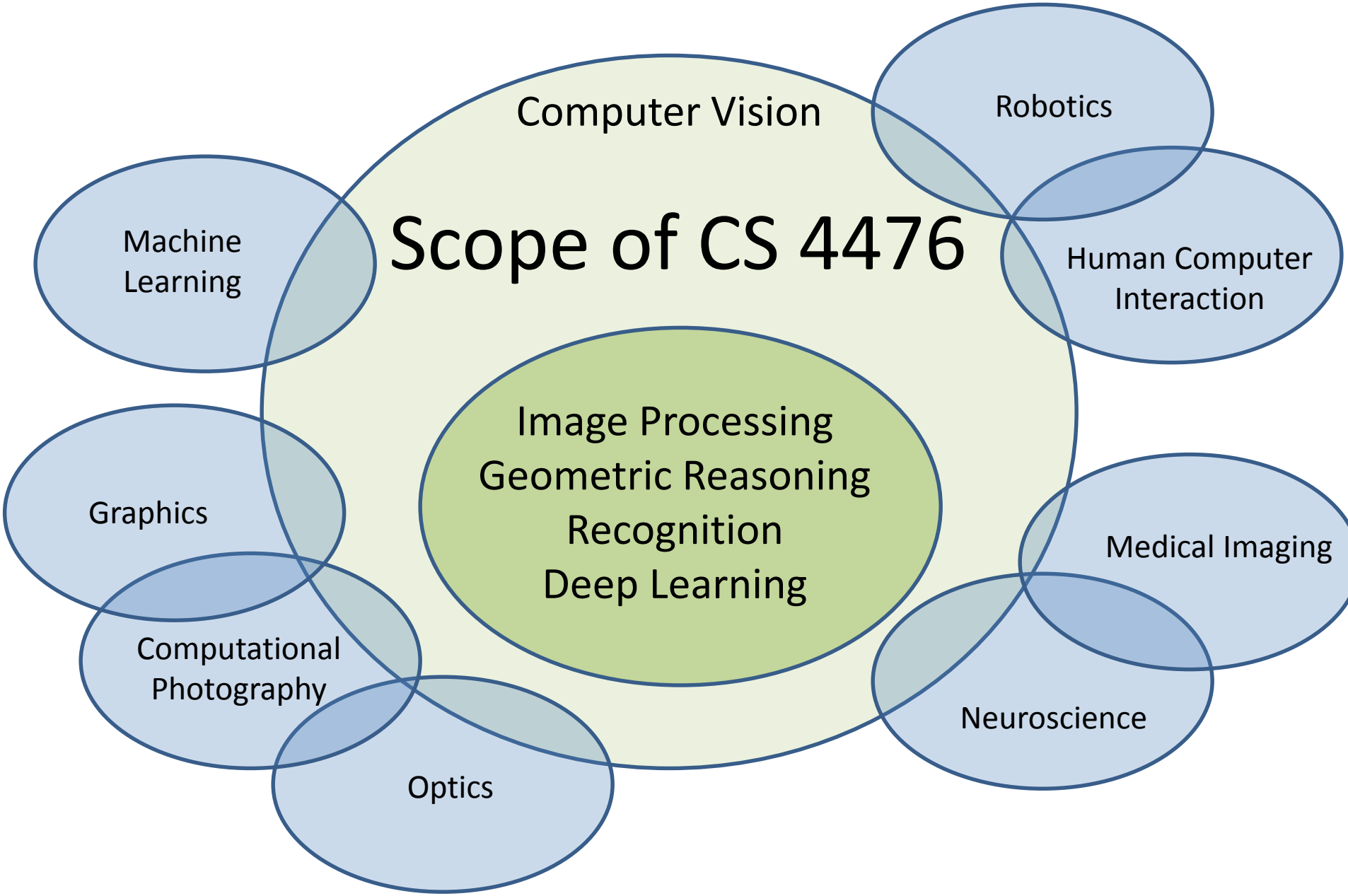
Machine learning applied to visual data

Course Syllabus (tentative)

http://www.cc.gatech.edu/~hays/compvision

Grading

- 80% programming projects (5 total)
- 20% quizzes (2 total)


Computer Vision
Scope of CS 4476
Image Processing
Geometric Reasoning
Recognition
Deep Learning
Robotics
Human Computer Interaction
Machine Learning
Graphics
Computational Photography
Optics
Medical Imaging
Neuroscience

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?
- 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, human-in-the-loop vision…

Textbook

Computer Vision: Algorithms and Applications

© 2010 Richard Szeliski, Microsoft Research

http://szeliski.org/Book/

Prerequisites

- **Linear algebra**, basic calculus, and probability
- Experience with image processing or Matlab will help but is not necessary

Tentative Projects

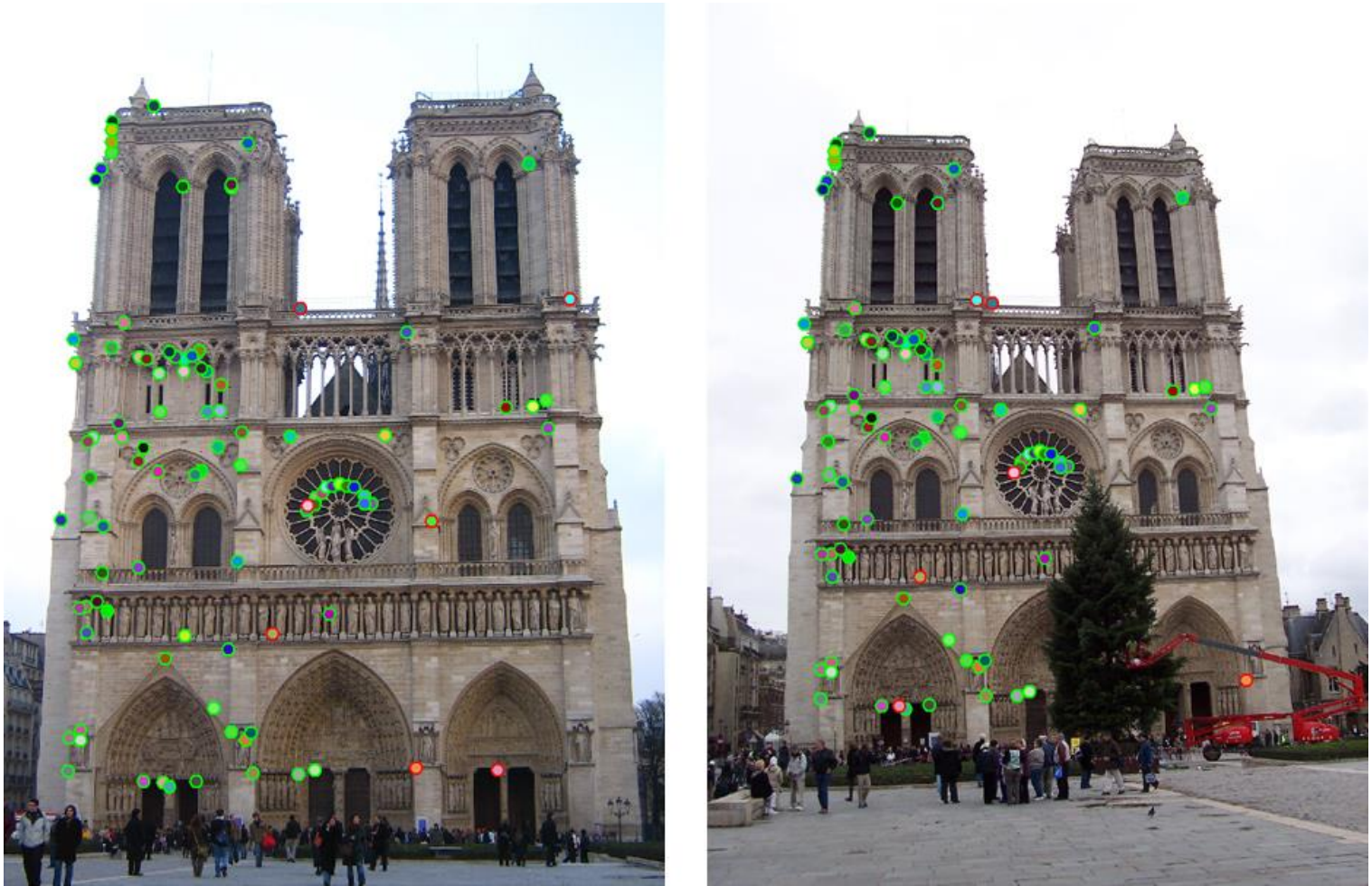
- Image Filtering and Hybrid Images
- Local Feature Matching
- Estimating Fundamental Matrices
- Scene Recognition with Bag of Words
- Object Detection with a Sliding Window
- Recognition with Deep Learning

Proj1: Image Filtering and Hybrid Images

- Implement image filtering to separate high and low frequencies
- Combine high frequencies and low frequencies from different images to create an image with scale-dependent interpretation

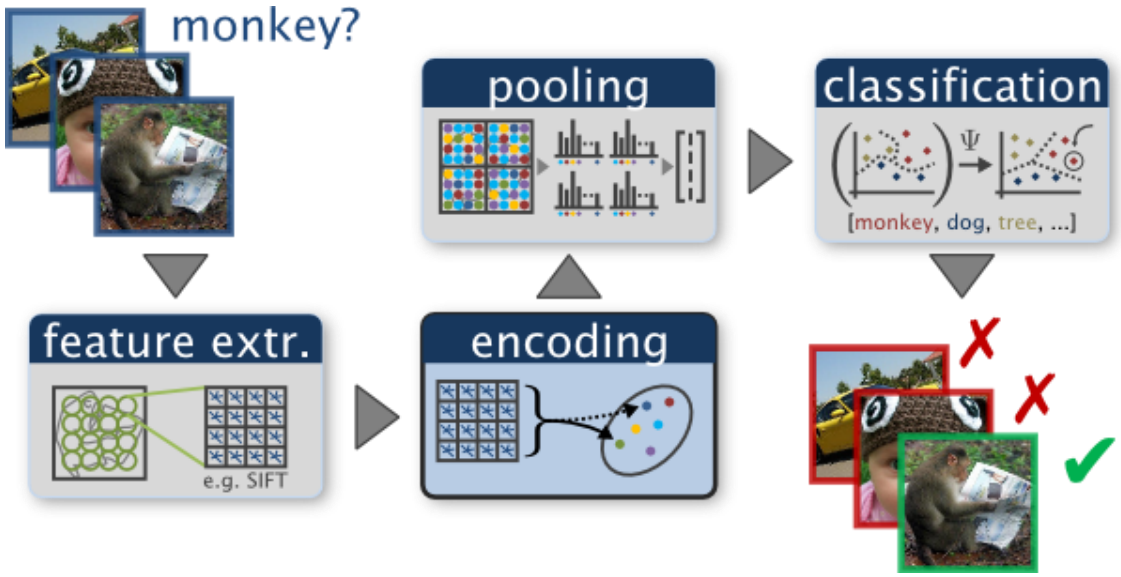
Proj2: Local Feature Matching

- Implement interest point detector, SIFT-like local feature descriptor, and simple matching algorithm.



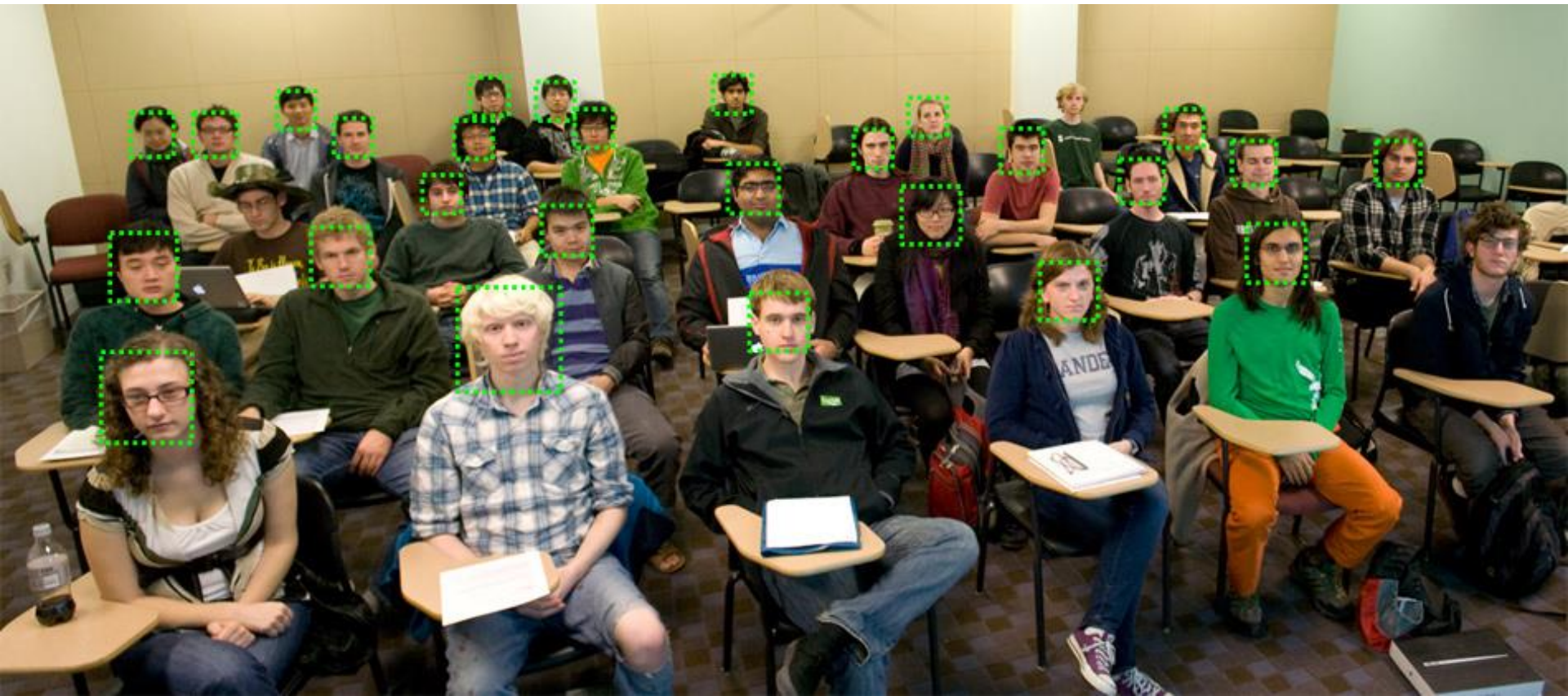
Proj4: Scene Recognition with Bag of Words

- Quantize local features into a “vocabulary”, describe images as histograms of “visual words”, train classifiers to recognize scenes based on these histograms.



Proj5: Object Detection with a Sliding Window

- Train a face detector based on positive examples and “mined” hard negatives, detect faces at multiple scales and suppress duplicate detections.



Course Syllabus (tentative)

http://www.cc.gatech.edu/~hays/compvision