CS 4476-A and 6476-A: Computer Vision

Instructor: James Hays
TAs: Otis Smith, Sooraj Karthik (head TAs), Mohit Aggarwal, Mansi Bhandari, SooHoon Choi, Deepanshi, Jesse Dill, Akhil Goel, Nikith Hosangadi, Haris Hussain, Jim James, Mark Kahoush, Xueqing Li, Alex Liu, Michael Propp, Aditya Sarma, Kelin Yu, Sili Zeng.
Today’s Class

• Who am I?
• What is Computer Vision?
• Specifics of this course
• Geometry of Image Formation
• Questions
A bit about me
What type of stuff do I work on?
Scene Flow from Point Clouds with or without Learning
Jhony Kaesemodel Pontes, James Hays, Simon Lucey
https://jhonykaesemodel.com/publication/sceneflow-3dv2020/
Understanding Lidar 3D for Free: Crossmodal Transfer Learning using HD Maps

Benjamin Wilson, Zsolt Kira, James Hays

Exploring new data sources

**ContactPose: A Dataset of Grasps with Object Contact and Hand Pose**

Samarth Brahmbhatt, Chengcheng Tang, Christopher D. Twigg, Charles C. Kemp, James Hays

ECCV 2020
Hand Contact Probability

handoff

use

Contact Probability
Patrick Grady, Chengcheng Tang, Samarth Brahmbhatt, Christopher D. Twigg, Chengde Wan, James Hays, and Charles C. Kemp

ECCV 2022 Oral
We train a deep network, PressureVisionNet, to estimate pressure from a single RGB image.

The pressure for each frame is calculated independently.
Exploring new data sources

https://www.argoverse.org/

**LIDAR**
- 2 roof-mounted LIDAR sensors
- Overlapping 40° vertical field of view
- Range of 200m
- On average, our LIDAR sensors produce a point cloud with ~107,000 points at 10 Hz

**Cameras**
- Seven high-resolution ring cameras (1920 x 1230) recording at 30 Hz with a combined 360° field of view
- Two front-view facing stereo cameras (2056 x 2464) sampled at 5 Hz

**Localization**
We use a city-specific coordinate system for vehicle localization. We include 6-DOF localization for each timestamp, from a combination of GPS-based and sensor-based localization methods.

**Calibration**
Sensor measurements for each driving session are stored in "logs." For each log, we provide intrinsic and extrinsic calibration data for LIDAR and all nine cameras.
Today’s Class

• Who am I?
• What is Computer Vision?
• Specifics of this course
• Geometry of Image Formation
• Questions
What is Computer Vision?

Derogatory summary of computer vision: Machine learning applied to visual data
Computer Vision

• Automatic understanding of images and video
  1. Computing properties of the 3D world from visual data (*measurement*)
1. Vision for measurement

Real-time stereo

Structure from motion

Tracking

Wang et al.

Snavely et al.

Demirdjian et al.

Slide credit: Kristen Grauman
Computer Vision

• Automatic understanding of images and video
  1. Computing properties of the 3D world from visual data (measurement)
  2. Algorithms and representations to allow a machine to recognize objects, people, scenes, and activities. (perception and interpretation)
2. Vision for perception, interpretation
Computer Vision

• Automatic understanding of images and video
  1. Computing properties of the 3D world from visual data (*measurement*)
  2. Algorithms and representations to allow a machine to recognize objects, people, scenes, and activities. (*perception and interpretation*)
  3. Algorithms to mine, search, and interact with visual data (*search and organization*)

Slide credit: Kristen Grauman
3. Visual search, organization

Slide credit: Kristen Grauman
Related disciplines

- Artificial intelligence
- Machine learning
- Cognitive science
- Computer vision
- Algorithms
- Image processing
- Graphics

Slide credit: Kristen Grauman
Vision and graphics

Inverse problems: analysis and synthesis.

Slide credit: Kristen Grauman
What humans see
What computers see
What do humans see?

Slide credit: Larry Zitnick
Vision is really hard

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

Is that a queen or a bishop?
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
• 2020’s: Widespread autonomous vehicles?
• 2030’s: robot uprising?
How vision is used now

• Examples of real-world applications
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
Optical character recognition (OCR)

• Most US postal service mail is automatically read.
• In 1997, there were 55 offices reviewing images of 19 billion pieces of mail that OCR failed on.
• Today, there is 1 office, and they only looked at 1.2 billion pieces of mail this year.

https://www.youtube.com/watch?v=XxCha4Kez9c
Face detection

• Digital cameras detect faces
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.
iNaturalist

https://www.inaturalist.org/pages/computer_vision_demo
Amazon Prime Air

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

https://www.skydio.com/
Given the camera feeds, we compute several important outputs on

https://www.youtube.com/watch?v=BVRMh9NO9Cs
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".
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*
Today’s Class

• Who am I?
• What is Computer Vision?
• Specifics of this course
• Geometry of Image Formation
• Questions
Grading

- 80% programming projects (5 total + 1 extra credit, maybe)
- 20% Quizzes or Problem sets

- Students in 6476 will have to do more for each project.

- We will have no final exam. The last project might extend into the final exam period.
Scope of CS 4476

Computer Vision

Image Processing
Geometric Reasoning
Recognition
Deep Learning

Robotics
Human Computer Interaction
Medical Imaging
Neuroscience
Optics
Computational Photography
Graphics
Machine Learning
Textbook


© 2020 Richard Szeliski, Facebook

http://szeliski.org/Book/
Prerequisites

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

You need a decent computer
You may want to buy a month of Google Colab Pro near the end of the semester
Projects

- (project 0 to test environment setup and handin)
- Image Filtering and Hybrid Images
- Local Feature Matching
- Camera Calibration and Fundamental Matrix Estimation with RANSAC
- Image Classification with Deep Learning
- Semantic Segmentation with Deep Learning
- Possibly a new extra credit project
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.
Course Syllabus (tentative)

https://faculty.cc.gatech.edu/~hays/compvision/
Code of Conduct

Your work must be your own. We’ll look for cheating. Don’t talk at the level of code with other students.
Today’s Class

- Who am I?
- What is Computer Vision?
- Specifics of this course
- Geometry of Image Formation
- Questions
The Geometry of Image Formation

Mapping between image and world coordinates
- Pinhole camera model
- Projective geometry
  - Vanishing points and lines
- Projection matrix
What do you need to make a camera from scratch?
Let’s design a camera
  – Idea 1: put a piece of film in front of an object
  – Do we get a reasonable image?

Slide source: Seitz
Pinhole camera

Idea 2: add a barrier to block off most of the rays

– This reduces blurring
– The opening known as the aperture

Slide source: Seitz
Pinhole camera

\[ f = \text{focal length} \]
\[ c = \text{center of the camera} \]
Camera obscura: the pre-camera

• Known during classical period in China and Greece (e.g. Mo-Ti, China, 470BC to 390BC)
Camera Obscursa used for Tracing

Lens Based Camera Obscura, 1568
Accidental Cameras

Accidental Pinhole and Pinspeck Cameras
Revealing the scene outside the picture.
Antonio Torralba, William T. Freeman
Accidental Cameras

a) Input (occluder present)
b) Reference (occluder absent)
c) Difference image (b-a)
d) Crop upside down

e) True view
First Photograph

Oldest surviving photograph
- Took 8 hours on pewter plate

Photograph of the first photograph

Joseph Niepce, 1826

Stored at UT Austin

Niepce later teamed up with Daguerre, who eventually created Daguerrotypes
“Louis Daguerre—the inventor of daguerreotype—shot what is not only the world's oldest photograph of Paris, but also the first photo with humans. The 10-minute long exposure was taken in 1839 in Place de la République and it's just possible to make out two blurry figures in the left-hand corner.”
Great history lesson on the chemistry and engineering challenges of early photography from the “Technology Connections” YouTube channel.

https://www.youtube.com/watch?v=wbbH77rYaa8&list=PLv0jwu7G_DFV6yW240e6CbiwCLaZ0Z6PV
Camera and World Geometry

How tall is this woman?

How high is the camera?

What is the camera rotation?

What is the focal length of the camera?

Which ball is closer?
Dimensionality Reduction Machine (3D to 2D)

3D world

Point of observation

2D image
Projection can be tricky...
Projection can be tricky...
Projective Geometry

What is lost?

• Length
Length and area are not preserved
Projective Geometry

What is lost?

• Length
• Angles
Projective Geometry

What is preserved?

• Straight lines are still straight
Vanishing points and lines

Parallel lines in the world intersect in the image at a “vanishing point”
Vanishing points and lines

Vanishing Point

Vanishing Line

Vanishing Point
Vanishing points and lines

Vertical vanishing point (at infinity)

Slide from Efros, Photo from Criminisi
Projection: world coordinates $\rightarrow$ image coordinates

If $x = 2$, $y = 3$, $z = 5$, and $f = 2$
What are $u$ and $v$?

$$u' = -x \cdot \frac{f}{z}$$  $$u' = -2 \cdot \frac{2}{5}$$

$$v' = -y \cdot \frac{f}{z}$$  $$v' = -3 \cdot \frac{2}{5}$$
Projection: world coordinates $\rightarrow$ image coordinates

How do we handle the general case?