When a user takes a photo, the app should check whether they're in a national park...

Sure, easy GIS lookup. Gimme a few hours.

...and check whether the photo is of a bird.

I'll need a research team and five years.
Plan for today

• Course Introduction
• Logistics
• Getting to know one another
The goal of computer vision

• To extract “meaning” from pixels

What we see

What a computer sees

Source: S. Narasimhan
What kind of information can be extracted from an image?

Source: L. Lazebnik
What kind of information can be extracted from an image?

**Geometric information**

Source: L. Lazebnik
What kind of information can be extracted from an image?

![Image with labeled objects]

**Geometric** information

**Semantic** information

Source: L. Lazebnik
Vision is easy for humans

Source: L. Lazebnik

Source: “80 million tiny images” by Torralba et al.
Vision is easy for humans

• Attneave’s Cat

Source: B. Hariharan
Vision is easy for humans

• Mooney Faces

Source: B. Hariharan
Vision is easy for humans

Surface perception in pictures. Koenderink, van Doorn and Kappers, 1992

Source: J. Malik
Remarkably Hard for Computers

When a user takes a photo, the app should check whether they're in a national park...

Sure, easy GIS lookup. Gimme a few hours.

...and check whether the photo is of a bird.

I'll need a research team and five years.

Source: XKCD
Vision is hard: Objects Blend Together

Source: B. Hariharan
Vision is hard: Objects Blend Together

Source: B. Hariharan
Vision is hard: Intra-class Variation

Source: B. Hariharan
Vision is hard: Intra-class Variation

Shape variation

Occlusion

Background clutter

Source: B. Hariharan
Vision is hard: Intra-class Variation
Vision is hard: Concepts are subtle

Tennessee Warbler

Orange Crowned Warbler

Source: B. Hariharan

https://www.allaboutbirds.org
Vision is hard: Images are ambiguous

Source: B. Hariharan
What can computer vision do today?
Reconstruction: 3D from photo collections


See also: [NYTimes Article](#)

Source: L. Lazebnik
Reconstruction: 4D from depth cameras

Figure 1: Real-time reconstructions of a moving scene with DynamicFusion; both the person and the camera are moving. The initially noisy and incomplete model is progressively denoised and completed over time (left to right).


[YouTube Video](#) Also see: *NeRF*

Source: L. Lazebnik
Reconstruction in construction industry

RECONSTRUCT INTEGRATES REALITY AND PLAN

**Visual Asset Management**

Reconstruct 4D point clouds and organize images and videos from smartphones, time-lapse cameras, and drones around the project schedule. View, annotate, and share anywhere with a web interface.

**4D Visual Production Models**

Integrate 4D point clouds with 4D BIM, review "who does what work at what location" on a daily basis and improve coordination and communication among project teams.

**Predictive Visual Data Analytics**

Analyze actual progress deviations by comparing Reality and Plan and predict risk with respect to the execution of the look-ahead schedule for each project location, to offer your project team with an opportunity to tap off potential delays before they surface on your jobsite.

reconstructinc.com
Applications

Source: N. Snavely
Recognition: “Simple” patterns

Source: L. Lazebnik
Recognition: Faces

Source: L. Lazebnik
Recognition: General categories

- **Computer Eyesight Gets a Lot More Accurate**, NY Times Bits blog, August 18, 2014
- **Building A Deeper Understanding of Images**, Google Research Blog, September 5, 2014

Source: L. Lazebnik
Recognition: General categories

- **ImageNet challenge**

- See also: **CLIP**

Source: L. Lazebnik
Object detection, instance segmentation

K. He, G. Gkioxari, P. Dollar, and R. Girshick, **Mask R-CNN**, ICCV 2017 (Best Paper Award)

Source: L. Lazebnik
Image generation

- Faces: 1024x1024 resolution, CelebA-HQ dataset


Source: L. Lazebnik | Follow-up work, NYTimes Article, DALL-E
Image generation

• BigGAN: 512 x 512 resolution, ImageNet

Easy classes

Difficult classes

A. Brock, J. Donahue, K. Simonyan, Large scale GAN training for high fidelity natural image synthesis, arXiv 2018

Source: L. Lazebnik
Origins of computer vision

L. G. Roberts *Machine Perception of Three Dimensional Solids*

Source: L. Lazebnik
Origins of computer vision

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
PROJECT MAC

Artificial Intelligence Group
Vision Memo. No. 100.

July 7, 1966

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".
Six decades of computer vision

1960s: Beginnings in artificial intelligence, image processing and pattern recognition

1970s: Foundational work on image formation: Horn, Koenderink, Longuet-Higgins ...

1980s: Vision as applied mathematics: geometry, multi-scale analysis, probabilistic modeling, control theory, optimization

1990s: Geometric analysis largely completed, vision meets graphics, statistical learning approaches resurface

2000s: Significant advances in visual recognition

2010s: Progress continues, aided by the availability of large amounts of visual data and massive computing power. Deep learning has become pre-eminent

Source: J. Malik
Growth of the field (attendance)

Long list of corporate sponsors
Growth of the field

CVPR Submitted and Accepted Papers

56% yearly growth with 26% acceleration → 10.8B submitted papers in 2028
Course overview

I. Early vision: Image formation and processing
II. Mid-level vision: Grouping and fitting
III. Multi-view geometry
IV. Recognition
V. Additional topics
I. Early vision

Basic image formation and processing

Cameras and sensors
Light and color

Feature extraction

Linear filtering
Edge detection

Optical flow

Source: L. Lazebnik
II. “Mid-level vision”

Fitting and grouping

Fitting: Least squares
Voting methods

Alignment

Source: L. Lazebnik
III. Multi-view geometry

Epipolar geometry

Two-view stereo

Structure from motion

Multi-view stereo

Source: L. Lazebnik
IV. Recognition

Basic classification

Deep learning

Object detection

Segmentation

Source: L. Lazebnik
V. Additional Topics (time permitting)

- Video
- 3D Scene Understanding
- Vision and Robotics

Source: L. Lazebnik
Logistics

• Course TAs

Wilfredo Calderon  Bowen Cheng  Amir Ibrahim

• Class website:
  http://saurabhg.web.illinois.edu/teaching/ece549/sp2021/