CAP5415
Computer Vision

Yogesh S Rawat
yogesh@ucf.edu
HEC-241
Course website

• https://www.crcv.ucf.edu/courses/cap5415-fall-2020/
Course logistics

• Class time: Tuesday and Thursday 4:30pm-5:45pm
  • Remote instruction: Zoom

• Office hours: Tuesday 2:00pm-3:00 pm
  • Zoom

• TA: Aayush Rana

• Office hours: Tuesday 3:00pm-4:00 pm
  • Zoom
Course outline

• Key topics
  • Mathematical Preliminaries
  • Imaging Geometry, Camera models, Coordinate Transforms
  • Image Filtering, Edge Detection, Feature Extraction
  • Basics of Neural Networks for Pattern Recognition
  • Deep Learning for Computer Vision
  • Region/Boundary Segmentation
  • Image Classification
  • Object Detection
  • Action Recognition
Course material

• Textbook
  • There is no required textbook for this course

• Suggested reference books are
  • Szeliski, Computer Vision: Algorithms and Applications (available online)
  • Shah, Fundamentals of Computer Vision (available from the course webpage)
  • Goodfellow et al. Deep Learning (available online)

• Pre-requisites
  • Basic Probability/Statistics
  • A good working knowledge of any programming language (python, C, etc.)
  • Linear algebra, Vector calculus.
Course logistics (requirements)

Exams:
  • No final exam
  • 2 Midterms

Homework
  • 10

Hands-on Experience:
  • 1 Final Project
  • 2 Programming assignments

Presentation Experience:
  • Final project presentation
Grading

• Homework
  • 1x10 = 10%

• Programming assignments
  • Two assignments, 2x10 = 20%

• Mid-term exams
  • Two written exams, 2x20 = 40%

• Project
  • Course project, 30% (code (15%), report(8%), presentation(7%))
  • Groups allowed, maximum 4 students
Grading criteria

• Grading
  • Programming assignments and the term project should include explanatory/clear comments as well as a short report describing the approach, detailed analysis, and discussion/conclusion.

• Programming
  • Students can use any programming language and platform of choice.

• Collaboration
  • Students are free to discuss ideas and technical concepts. However, students must submit original work for all assignments, projects and exams, and abide by UCF Golden Rule.
Course expectations

• Attend class lectures
• Timely submission of homework
  • 0 for late submission
• Homework questions will be simple
  • From class lectures
  • True/False, fill in the blanks, multiple choice, 5 questions
  • Mostly every week, total 10
  • Release on Thursday, due on Monday
• Programming assignments
  • 2 small coding assignments
Course expectations

• Mid-terms
  • Questions from class lectures
  • First mid-term for first half
  • Second mid-term for second half

• Course project
  • Select from a list we will provide
  • or your own choice
  • Grading will be based on
    • Submitted code
    • Report
    • Final presentations
<table>
<thead>
<tr>
<th>Lecture</th>
<th>Date</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lecture 1</td>
<td>Tue, 08/25/2020</td>
<td>Introduction to Computer Vision</td>
</tr>
<tr>
<td>Lecture 2</td>
<td>Thu, 08/27/2020</td>
<td>Linear Algebra</td>
</tr>
<tr>
<td>Lecture 3</td>
<td>Tue, 09/01/2020</td>
<td>Image Filtering</td>
</tr>
<tr>
<td>Lecture 4</td>
<td>Tue, 09/08/2020</td>
<td>Introduction to Deep Learning</td>
</tr>
<tr>
<td>Lecture 5</td>
<td>Tue, 09/15/2020</td>
<td>PyTorch Tutorial</td>
</tr>
<tr>
<td>Lecture 6</td>
<td>Thu, 09/24/2020</td>
<td>Edge Detection</td>
</tr>
<tr>
<td>Lecture 7</td>
<td>Thu, 10/01/2020</td>
<td>Interest Point Detection</td>
</tr>
<tr>
<td>Lecture 8</td>
<td>Thu, 10/08/2020</td>
<td>Features</td>
</tr>
<tr>
<td>Lecture 9</td>
<td>Thu, 10/15/2020</td>
<td>Classification</td>
</tr>
<tr>
<td></td>
<td>Thu, 10/22/2020</td>
<td><strong>Mid-Term Exam - I</strong></td>
</tr>
<tr>
<td>Lecture 10</td>
<td>Thu, 10/29/2020</td>
<td>Object Detection</td>
</tr>
<tr>
<td>Lecture 11</td>
<td>Thu, 11/05/2020</td>
<td>Segmentation</td>
</tr>
<tr>
<td>Lecture 12</td>
<td>Thu, 11/12/2020</td>
<td>Optical Flow</td>
</tr>
<tr>
<td>Lecture 13</td>
<td>Tue, 11/17/2020</td>
<td>Action Recognition</td>
</tr>
<tr>
<td>Lecture 14</td>
<td>Tue, 11/24/2020</td>
<td>Camera Model</td>
</tr>
<tr>
<td></td>
<td>Tue, 12/01/2020</td>
<td><strong>Mid-Term Exam - II</strong></td>
</tr>
</tbody>
</table>

https://www.crcv.ucf.edu/courses/cap5415-fall-2020/schedule/
Questions?
Introduction to Computer Vision

Lecture 1
What is computer vision?

• Ability of computers
  • To understand visual data
    • For example, images, videos...

• Automate tasks
  • Which human visual system can perform
What is it related to?
Visual Perception

• How do we know that the objects that we see are for real?
• Can people “see” without being aware of what they see?
• Why do objects appear colored?
**Visual Perception**

- **Definition:** Process of acquiring knowledge about environmental objects and events by extracting information from the light they emit or reflect [Palmer, 2012].

Perception is analogous to taking a picture! (credit: Palmer, 2012)
**Visual Perception**

*Vision* is a process in which temporally changing intensity and color values in the image plane have to be interpreted as processes in the real world that happen in 3D space over time.
What can you see in this picture?
Now can you see what the picture is?

Vision vs. Computer Vision?

Sensing device → Interpreting device → Interpretations

Picture
Man
Thrash
Bulb
Light
...

Lecture 1 - Introduction
Vision and Image Understanding

• **Visual tasks:** We use vision to interact with environments and survive
  • to navigate and avoid obstacles
  • to recognize and pick up objects
  • to identify food and danger
  • friends and enemies
  • ...

Lecture 1 - Introduction
Goal of Computer Vision?

- To bridge the gap between
  - image pixels and “meaning” (semantic)!

What we see!

What computer sees!
What is a (digital) Image?

• Definition: A digital image is defined by *integrating* and *sampling* continuous (analog) data in a spatial domain [Klette, 2014].
Image Types: (Gray)Scalar and Binary

• A scalar image has integer values

\[ u \in \{0, 1, \ldots, 2^a - 1\} \]

• a: level (bit)

• **Ex.** If 8 bit (a=8)
  • image spans from 0 to 255
  • 0 black and 255 white

• **Ex.** If 1 bit (a=1)
  • it is binary image
  • 0 and 1 only

![](image.png)
Image Type: RGB (red, green, blue)

- Image has three channels (bands)
- Each channel spans a-bit values.

Human Cone-cells (normalized) responsivity spectra

- Some people might have 4 cone-types!
- Some might have just 2!
Color

• Color vision has evolved over millions of years.
If there is **no light**, there is **no color**!

Human vision can only discriminate a few dozens of grey levels on a screen, but hundreds of thousands of different colors.

- **RED** -> ~625 to 780 nm  [long wavelength]
- **ORANGE** -> ~590 to 625 nm  [long wavelength]
- **YELLOW** -> ~565 to 590 nm  [middle range wavelength]
- **GREEN** -> ~500 to 565 nm  [middle range wavelength]
- **CYAN** -> ~485 to 500 nm  [middle range wavelength]
- **BLUE** -> ~440 to 485 nm  [short wavelength]
- **VIOLET** -> ~330 to 440 nm  [very short wavelength]
There are three different types of color-sensitive cones corresponding to (roughly)

- RED (64% of the cones)
- GREEN (about 32%), and
- BLUE (about 2%).

6-7 million cones
120 million rods
Motivation

Why study computer vision?
Because...

• Find it interesting!
• Out of curiosity!
• Course requirement!
Recent advancement

• Used to be done mostly in academics

• Now
  • Google
  • Facebook
  • Apple
  • IBM
  • Microsoft
  • ...

8/25/2020
What changed?

• Emergence of deep learning
• Advancement in hardware
• Availability of large-scale data
  • ImageNet
  • OpenImages
  • YFCC100M
  • Youtube-8M
  • Kinetics
  • AVA
  • …
Investment in computer vision

The graph illustrates the investment trends in computer vision (CV) companies in the United States and China from 2011 to 2017. It shows the total capital invested in CV companies in millions and the number of deals in these companies. The data indicates a significant increase in investment, particularly in China, during this period.
# CVPR conference ranking (Engineering)

<table>
<thead>
<tr>
<th>Publication</th>
<th>h5-index</th>
<th>h5-median</th>
</tr>
</thead>
<tbody>
<tr>
<td>IEEE/CVF Conference on Computer Vision and Pattern Recognition</td>
<td>299</td>
<td>509</td>
</tr>
<tr>
<td>Advanced Materials</td>
<td>273</td>
<td>369</td>
</tr>
<tr>
<td>International Conference on Learning Representations</td>
<td>203</td>
<td>359</td>
</tr>
<tr>
<td>Renewable and Sustainable Energy Reviews</td>
<td>201</td>
<td>263</td>
</tr>
</tbody>
</table>
CVPR attendance
Police chase

https://www.youtube.com/watch?v=VMxjFTtTquc
Applications
Retail – Amazon Go

https://youtu.be/NrmMk1Myrxc
Retail - StopLift

https://youtu.be/pBpnC9QvGm4
Retail - Clothing

https://youtu.be/Mr71jrkwWq8
Self-driving - Waymo

https://youtu.be/TsaES--OTzM
Autopilot - Tesla

https://youtu.be/UgNhYGAgmZo
Computer vision tasks
Object Recognition

- **Problem:** Given an image A, does A contain an image of a person?
Object Recognition

• **Problem:** Given an image A, does A contain an image of a person?
Object Recognition

• **Problem:** Given an image A, does A contain an image of a person?
Object localization
Human Detection
Semantic Segmentation

Image source: Antonio Torralba
Segmentation

Image source: Antonio Torralba
Segmentation

Image source: Antonio Torralba
Segmentation

PEOPLE WALKING IN THE PARK

PERSON FEEDING DUCKS IN THE PARK

DUCKS LOOKING FOR FOOD

Do not feed the ducks sign

Image source: Antonio Torralba
Semantic Segmentation: Results

Semantic part labeling
Face Recognition
Open-Universe Face Identification

News Article: Label Important Figures

Social Network: Tag Facebook Friends
Open-Universe Face Identification

Find Angelina Jolie and George Clooney
Open-Universe Face Identification

Find Angelina Jolie and George Clooney
Facial expression
Fatigue detection
Lip-reading
High Density Crowded Scenes

- Political Rallies
- Religious Festivals
- Marathons
- High Density Moving Objects
Counting

Ground truth=634  Proposed Method by Idrees and Shah=640
Counting

Ground truth=1428, Proposed Method=1468
Counting

Ground truth=2319   Proposed Method=2496
Visual Business Recognition

NAME: Pizza My Heart
ADDRESS: 220 University Ave, Palo Alto, CA 94301
USER Rating: 3.5/5
CATEGORY: Pizza
PHONE: (650) 327-9400
Biometrics

Fingerprint scanners on many new laptops, other devices

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

Source: S. Seitz
Smile detection

The Smile Shutter flow

Imagine a camera smart enough to catch every smile! In Smile Shutter Mode, your Cyber-shot® camera can automatically trip the shutter at just the right instant to catch the perfect expression.

Source: S. Seitz
Vision-based interaction (and games)

Microsoft’s Kinect

Sony EyeToy

Assistive technologies

Source: S. Seitz
Video Clip

• Sequences of frames
• 30 frames per second
Sequences of Images
Action recognition – UCF101

- Cycling
- Diving
- Golf Swinging
- Riding
- Volleyball
- Basketball Shooting
- Swinging
- Tennis Swinging
Action detection
Video segmentation

Duarte, Rawat, Shah, ICCV 2019
Cross-view action synthesis

ECCV 2020
Detection in aerial videos
(Object) Tracking
Tracking (multi-object)
Video Surveillance and Monitoring

Object detection → Object tracking → Object categorization and classification → Event or Activities Recognition
So…. How do we detect an object in an image?
Naïve approach: Template Matching

Find the chair in this image

Output of correlation

This is a chair
Template Matching

Find the chair in this image

Epic fail!
Simple template matching is not going to make it

Image source: Antonio Torralba
Idea

• Instead of comparing raw image pixels:
  • First map those pixels into another (more robust) form,
  • And then compare those mapped forms.
  • Finally, select the closest image map (how do you define “closest”? Metrics).

• Features, examples:
  • compute edges
  • compute color histograms
  • gradients
  • HOG
  • SIFT
  • ...
“Bag-of-Words” Representation

Image source: Antonio Torralba
“Bag-of-Words” (BoW) Histograms

Image Features → BoW Histogram → Classifier → Result

Recipe
BoW Representation

- All have equal probability for bag-of-words methods,
- Location (spatial) information is important but lost.

Image source: Antonio Torralba
Computer vs. Human Vision?

• In which task computers are superior to humans?
• In which task humans are superior to computers?
Human > Computer
Recognition, detection, etc ... global tasks

Computer > Human
Delineation, local analysis, etc ... local tasks
A brief history
50 Years Ago

Out of memory
25 Years ago
15 Years ago
5 years ago
Today

https://youtu.be/-QvIX3cY4lc
Questions?

Sources for this lecture include materials from works by Abhijit Mahalanobis, Sedat Ozer, Ulas Bagci, Mubarak Shah, and Antonio Torralba