UCF Center for Research in Computer Vision

CAP5415 Computer Vision

Yogesh S Rawat

yogesh@ucf.edu

HEC-241

All lectures will be recorded!

Administrative details

Course logistics

- Class time:
  - Tuesday and Thursday
  - Time: 4:30pm-5:45pm
  - In-person: BA1 O146
  - Remote instruction: Zoom
- Office hours:
  - Tuesday: 2:00 pm-3:00 pm
  - Thursday: 3:00 pm-4:00 pm
  - Office HEC241 and Zoom
- TA: Binay Singh
  - Office hours: Tuesday 3:00pm-4:00 pm, Thursday 2:00pm-3:00 pm
  - ENG2 237 and Zoom

Course outline

- Key topics
  - Mathematical Preliminaries
  - Imaging Geometry, 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
  - Some recent trend: NeRF, Transformers, Diffusion Models (Overview)

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 on course webpage**)
  - Goodfellow et al. Deep Learning (**available online**)
- Pre-requisites
  - Basic Probability/Statistics
  - A good working knowledge of Python [**this is really important**]
  - Linear algebra, Vector calculus.

Course logistics (requirements)

Exams:

- Midterm
- Final exam

Homework

- 10

Hands-on experience:

- 4 Programming assignments
- 1 Final Project

Grading

- Homework
  - 1x10 = 10%
- Programming assignments
  - Four assignments, 4x5 = 20%
- Mid-term/Final exams
  - Two written exams, 2x20 = 40%
- Project
  - Course project, 30% (code (10%), report(20%))
  - Groups allowed, maximum 6 students

Grading

- BONUS
  - Pop Quiz
  - 1 programming assignment
    - NOTE: only towards programming assignment

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
  - Python preferred
  - 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
  - Watch recorded lectures before class
  - Come prepared with questions
- 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
  - 4 coding assignments

Course expectations

- Mid-term/Final exam
  - Questions from class lectures
  - Mid-term for first half
  - Final exam for second half
- Course project
  - Select from a list we will provide
  - or your own choice
  - Grading will be based on
    - Submitted code
    - Report

Schedule

Lecture 1	Tue, 08/22/2023	Introduction to Computer Vision
Lecture 2	Thu, 08/24/2023	Linear Algebra
Lecture 3	Tue, 08/29/2023	Image Filtering
Lecture 4	Tue, 09/05/2023	Introduction to Deep Learning
Lecture 5	Tue, 09/12/2023	PyTorch Tutorial
Lecture 6	Thu, 09/21/2023	Edge Detection
Lecture 7	Thu, 09/28/2023	Interest Point Detection
Lecture 8	Thu, 10/06/2023	Features
Lecture 9	Thu, 10/10/2023	Classification
	Thu, 10/18/2023	**Mid-Term Exam**
Lecture 10	Thu, 10/24/2023	Object Detection
Lecture 11	Thu, 11/07/2023	Segmentation
Lecture 12	Thu, 11/16/2023	Optical Flow
Lecture 13	Tue, 11/21/2023	Action Recognition
Lecture 14	Tue, 11/28/2023	Advance topics
	Thu, 12/07/2023	**Final Exam**

https://www.crcv.ucf.edu/courses/cap5415-fall-2023/schedule/

Feedback

- During the course
- A link was shared via Webcourses

Homework – 0

- Due on Friday
- No other homework this week
- Homework – 1
  - Will be released next Thursday

Course website

- https://www.crcv.ucf.edu/courses/cap5415-fall-2022/

Questions?

Motivation

Recent advances

Instance segmentation

https://github.com/IDEA-Research/Grounded-Segment-Anything

Panoptic segmentation

https://github.com/IDEA-Research/Grounded-Segment-Anything

Manipulation

https://github.com/IDEA-Research/Grounded-Segment-Anything

Mesh recovery

https://github.com/IDEA-Research/OSX

Questions?

CAP5415 Computer Vision

Yogesh S Rawat

yogesh@ucf.edu

HEC-241

Introduction to Computer Vision

Lecture 1

Introduction to Computer Vision

Lecture 1

Basics

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].*

Vision → Cognitive Activity → Acquisition of knowledge

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?

Credit: Thompson, Basic Vision, Oxford Press, 2012.

Vision vs. Computer Vision?

Sensing device

Interpreting device

Interpretations

Picture
Man
Thrash
Bulb
Light
…

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
  - …

Goal of Computer Vision?

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

What we see!

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

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].

Left hand coordinate system

Image Types: (Gray)Scalar and Binary

- A scalar image has integer values

$$u \in \{0, 1, ..., 2^a - 1\}$$

- 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

Original (256 colors) | 8 colors | 4 colors

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.

Color

- If there is no light, there is no color!
- Human vision can only discriminate a few dozens of grey levels
  - 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]

Retina of Human Eye

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 [for colors]
120 million rods [for light]

Some may have only 2 cones
Dogs also have 2 types

Credit: Klette, 2012.

Questions?

Introduction to Computer Vision

Lecture 1

Motivation

Motivation

Why study computer vision?

Because…

- Find it interesting!
- Out of curiosity!
- Course requirement!
- Good job market!

Recent advancement

- Used to be done mostly in academics
- Now
  - Google
  - Facebook
  - Apple
  - IBM
  - Microsoft
  - …

What changed?

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

Google Cloud

Microsoft Azure

NVIDIA GPU CLOUD

amazon web services™

Market of computer vision

Source: https://www.iflexion.com/blog/computer-vision-applications

CVPR conference ranking (Engineering)

	Publication	h5-index	h5-median
1.	IEEE/CVF Conference on Computer Vision and Pattern Recognition	389	627
2.	Advanced Materials	312	418
3.	International Conference on Learning Representations	286	533
4.	Neural Information Processing Systems	278	436

CVPR conference ranking

	Publication	h5-index	h5-median
1.	Nature	444	667
2.	The New England Journal of Medicine	432	780
3.	Science	401	614
4.	IEEE/CVF Conference on Computer Vision and Pattern Recognition	389	627

CVPR - papers

https://medium.com/reconstruct-inc/the-golden-age-of-computer-vision-338da3e471d1

CVPR attendance

CVPR 2022 in Numbers

8,161 Paper submissions

2,064 Papers accepted

7,289 Registered attendees (05/31/2022)
5510 in person, 1779 virtual

9,981 Registered attendees (06/20/2022)
5641 in person, 4340 virtual

Questions?

Introduction to Computer Vision

Lecture 1

Applications

Applications

Retail – Amazon Go

https://youtu.be/NrmMk1Myrxc

Retail - Clothing

https://youtu.be/Mr71jrkzWq8

Self-driving - Waymo

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

Autopilot - Tesla

https://youtu.be/UgNhYGAgmZo

Questions?

Introduction to Computer Vision

Lecture 1

Tasks

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?

YES

Object Recognition

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

NO

Object localization

Human Detection

Semantic Segmentation

Image source: Antonio Torralba

Segmentation

Image source: Antonio Torralba

Segmentation

Image source: Antonio Torralba

Segmentation

Image source: Antonio Torralba

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

Smile detection

Sony Cyber-shot® T70 Digital Still Camera

Source: S. Seitz

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=2319 Proposed Method=2496

Visual Business Recognition

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 Security 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

This is a chair

Find the chair in this image

Output of correlation

Image source: Antonio Torralba

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

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

25 Years ago

15 Years ago

5 years ago

Today

https://youtu.be/PINeQV0LH6k

Is it perfect?

Stop Sign

Speed 60

Questions?

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