CSC420: Intro to Image Understanding

Introduction

Ahmed Ashraf

September 7, 2017
(Content inherited from Sanja Fidler)
The Team

- **Instructor:**
  - Ahmed Ashraf (aashraf@alumni.cmu.edu)
  - Office: BA 3219
  - Office hours: Mondays and Fridays 4-5pm

- **TAs:**
  - Jake Snell (jsnell@cs.toronto.edu)
  - Mohammad Kianpisheh (mkian1368@gmail.com)
  - Hang Chu (chuhang1122@cs.toronto.edu)
  - Andrei Barsan (andrei.ioan.barsan@gmail.com)
Course Information

- **Class time**: Tuesday and Thursday at 3-4pm
- **Location**: GB248
- **Tutorials**: on demand
- **Class Website**: Coming soon

The class will use Piazza for **announcements and discussions**:

https://piazza.com/utoronto.ca/fall2017/csc420/home

- Your grade will **not depend on your participation on Piazza**. It’s just a good way for asking questions, discussing with your instructor, TAs and your peers
Course Information

- **Textbook**: We won’t directly follow any book, but extra reading in this textbook will be useful:

  Rick Szeliski

  *Computer Vision: Algorithms and Applications*

  available free online:

  [http://szeliski.org/Book/](http://szeliski.org/Book/)

- Links to other material (papers, code, etc) will be posted on the class webpage
Course Prerequisites:

- Data structures
- Linear Algebra
- Vector calculus

Without this you’ll need some serious catching up to do!

Knowing some basics in this is a plus:

- Matlab, Python, C++
- Machine Learning
- Neural Networks
- Solving assignments sooner rather than later
Requirements and Grading

- Each student expected to complete 5 assignments and a project

**Grading**

- **Assignments**: 60% (12% each)
- **Project**: 40%

**Assignments:**

- Short *theoretical questions* and *programming exercises*
- Will be given every *two weeks* (starting with second week of class)
- You will have *a week to hand in the solution* to each assignment
- You need to solve the assignment *alone*

**Project:**

- You will be able to choose from a list of projects or come up with your own project (discussed prior with your instructor)
- Need to hand in a *report* and do an oral *presentation*
- Can work in up to teams of 3
<table>
<thead>
<tr>
<th>Term Work</th>
<th>Post Date</th>
<th>Due Date</th>
<th>% of grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assignment 1</td>
<td>Sept 14</td>
<td>Sept 21</td>
<td>12%</td>
</tr>
<tr>
<td>Assignment 2</td>
<td>Sept 28</td>
<td>Oct 5</td>
<td>12%</td>
</tr>
<tr>
<td>Assignment 3</td>
<td>Oct 12</td>
<td>Oct 19</td>
<td>12%</td>
</tr>
<tr>
<td>Project Proposal</td>
<td></td>
<td>Oct 24</td>
<td>5%</td>
</tr>
<tr>
<td>Assignment 4</td>
<td>Oct 26</td>
<td>Nov 2</td>
<td>12%</td>
</tr>
<tr>
<td>Assignment 5</td>
<td>Nov 16</td>
<td>Nov 23</td>
<td>12%</td>
</tr>
<tr>
<td>Project Report</td>
<td></td>
<td>Nov 25</td>
<td>20%</td>
</tr>
<tr>
<td>Project Presentation</td>
<td>Nov 28,30,Dec 5</td>
<td></td>
<td>15%</td>
</tr>
</tbody>
</table>
Programming Language?

- Your assignments / project can be in Matlab, Python, C++
- As long as it compiles, runs, and you know how to defend it, we’re happy
- HOWEVER, most code and examples we will provide during the class will be in Matlab
- Choose wisely
Lateness

**Deadline** The solutions to the assignments / project should be submitted by **11.59pm on the date they are due**. Anything from 1 minute late to 24 hours will count as one late day.

**Lateness** Each student will be given a total of **3 free late days**. This means that you can hand in three of the assignments one day late, or one assignment three days late. It is up to you to make a good planning of your work. **After you have used the 3 day budget, the late assignments will not be accepted.**
# Tentative syllabus

<table>
<thead>
<tr>
<th>Week nb.</th>
<th>Date</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sept 7</td>
<td>Intro</td>
</tr>
<tr>
<td>2</td>
<td>Sept 12 &amp; Sept 14</td>
<td>Linear filters, edges</td>
</tr>
<tr>
<td>3</td>
<td>Sept 19 &amp; Sept 21</td>
<td>Image features</td>
</tr>
<tr>
<td>4</td>
<td>Sept 26 &amp; Sept 28</td>
<td>Keypoint detection</td>
</tr>
<tr>
<td>5</td>
<td>Oct 3 &amp; Oct 5</td>
<td>Matching</td>
</tr>
<tr>
<td>6</td>
<td>Oct 10 &amp; Oct 12</td>
<td>Segmentation</td>
</tr>
<tr>
<td>7</td>
<td>Oct 17 &amp; Oct 19</td>
<td>Grouping</td>
</tr>
<tr>
<td>8</td>
<td>Oct 24 &amp; Oct 26</td>
<td>Object recognition</td>
</tr>
<tr>
<td>9</td>
<td>Oct 31 &amp; Nov 2</td>
<td>Object detection</td>
</tr>
<tr>
<td>10</td>
<td>Nov 14 &amp; Nov 16</td>
<td>Neural Networks</td>
</tr>
<tr>
<td>11</td>
<td>Nov 21 &amp; Nov 23</td>
<td>Stereo, multi-view</td>
</tr>
<tr>
<td>12</td>
<td>Nov 25,28 &amp; Dec 5</td>
<td>Project Presentations</td>
</tr>
</tbody>
</table>
Introduction to Intro to Image Understanding

- What is Computer Vision?
- Why study Computer Vision?
- Which cool applications can we do with it?
- Is vision a hard problem?
What is Computer Vision?
What is Computer Vision?

- A field trying to develop automatic algorithms that would "see"
What is Computer Vision?

- What does it mean to see?
  - To know what is where by looking – Marr, 1982

[text adopted from A. Torralba]
What is Computer Vision?

- What does it mean to see?
  - To know what is where by looking – Marr, 1982
  - Understand where things are in the world
What is Computer Vision?

- What does it mean to see?
  - To know what is where by looking – Marr, 1982
  - Understand where things are in the world
  - What are their 3D properties?

[Text adopted from A. Torralba]
What is Computer Vision?

- What does it mean to see?
  - To know what is where by looking – Marr, 1982
  - Understand where things are in the world
  - What are their 3D properties?

Depth pic from http://vladlen.info
What is Computer Vision?

- What does it mean to see?
  - To know what is where by looking – Marr, 1982
  - Understand where things are in the world
  - What are their 3D properties?
  - What actions are taking place?

Pic from www.cobblehillpuzzles.com

Ahmed Ashraf
Intro to Image Understanding
“Full” Image Understanding?

- Full understanding of an image?
“Full” Image Understanding?

- Full understanding of an image? You can answer any question about it

“Full” Image Understanding?

- Full understanding of an image? You can answer any question about it

Q: What is behind the table? A: window

Q: What is in front of the toilet? A: door

Q: What is on the counter in the corner? A: microwave
“Full” Image Understanding?

- Full understanding of an image? **You can answer any question about it**

![Image of a dining room](image1.jpg)
Q: What is behind the table?  
A: window

![Image of a bathroom](image2.jpg)
Q: What is in front of the toilet?  
A: door

![Image of a kitchen](image3.jpg)
Q: What is on the counter in the corner?  
A: microwave

![Image of a classroom](image4.jpg)
Q: What is the shape of the green chair?  
A: horse shaped
“Full” Image Understanding?

- Full understanding of an image? **You can answer any question about it**

  ![Image 1](image1.png)  ![Image 2](image2.png)  ![Image 3](image3.png)

  Q: What is behind the table?  
  A: window

  Q: What is in front of the toilet?  
  A: door

  Q: What is on the counter in the corner?  
  A: microwave

  ![Image 4](image4.png)  ![Image 5](image5.png)

  Q: What is the shape of the green chair?  
  A: horse shaped

  Q: Where is the oven?  
  A: on the right side of the fridge
“Full” Image Understanding?

- Full understanding of an image? **You can answer any question about it**

  - Q: What is behind the table?  
    A: window

  - Q: What is in front of the toilet?  
    A: door

  - Q: What is on the counter in the corner?  
    A: microwave

  - Q: What is the shape of the green chair?  
    A: horse shaped

  - Q: Where is the oven?  
    A: on the right side of the fridge

  - Q: What is the largest object?  
    A: bed
“Full” Image Understanding?

- Full understanding of an image? **You can answer any question about it.**

Q: Which object is red?
A: toaster
Full understanding of an image? **You can answer any question about it**

1. **Q:** Which object is red?  
   **A:** toaster

2. **Q:** How many drawers are there?  
   **A:** 6

3. **Q:** How many doors are open?  
   **A:** 1

4. **Q:** How many lights are on?  
   **A:** 6
Full understanding of an image? **You can answer any question about it**

- **Q:** Which object is red?  
  **A:** toaster

- **Q:** How many drawers are there?  
  **A:** 6

- **Q:** How many doors are open  
  **A:** 1

- **Q:** How many lights are on?  
  **A:** 6

- **Q:** Can you make pizza in this room?  
  **A:** yes

- **Q:** Where can you sit?  
  **A:** chairs, table, floor
Why study Computer Vision?
Why study Computer Vision?

- Because you want your robot to fold your laundry
Why study Computer Vision?

- ... and drive you to work (video)

Amnon Shashua’s Mobileye autonomous driving system
Why study Computer Vision?

- Allows you to manipulate your images

*Scene Completion using Millions of Photographs*, Hays & Efros, SIGGRAPH 2007
Why study Computer Vision?

- Allows you to manipulate your images

*Scene Completion using Millions of Photographs*, Hays & Efros, SIGGRAPH 2007
Why study Computer Vision?

- Allows you to manipulate your images

*Scene Completion using Millions of Photographs, Hays & Efros, SIGGRAPH 2007*
Why study Computer Vision?

- Allows you to manipulate your images

*Scene Completion using Millions of Photographs*, Hays & Efros, SIGGRAPH 2007
Why study Computer Vision?

- Allows you to manipulate your images

*Scene Completion using Millions of Photographs*, Hays & Efros, SIGGRAPH 2007
Why study Computer Vision?

- Allows you to manipulate your images

*Scene Completion using Millions of Photographs, Hays & Efros, SIGGRAPH 2007*
Why study Computer Vision?

- Change style of images

[A] [B] [C] [D]

[Butler, Ecker, Bethge. A Neural Algorithm of Artistic Style. Arxiv'15.]
Why study Computer Vision?

- ... and make cool videos using a single image

3D Object Manipulation in a Single Photograph using Stock 3D Models, Kholgade, Simon, Efros, Sheikh, SIGGRAPH 2014
Why study Computer Vision?

- Reconstruct the world in 3D from online photos!

Photosynth, https://photosynth.net/ (try it!)
Why study Computer Vision?

- Figure out what people are wearing

http://clothingparsing.com (try it!)
Why study Computer Vision?

- Crazy media attention!!!
Why study Computer Vision?

- Crazy media attention!!!

From Cosmopolitan: *The technology scores your facial attributes (this just keeps getting better, doesn’t it) from your looks, to your age, and the emotion you’re showing, before combining all the information using an equation SO complex we won’t begin to go into it.*
Why study Computer Vision?

- Detect and analyze faces

http://www.rekognition.com (try it!)
Why study Computer Vision?

- Detect and analyze faces

http://www.rekognition.com  (try it!)
Why study Computer Vision?

- You can make yourself look better (and competitors worse)

[Khosla, Bainbridge, Oliva, Torralba, Modifying the Memorability of Face Photographs, ICCV 2013]
Why study Computer Vision?

- An Eye for Pain (video)

Ashraf et al’s Pain Detection System
Why study Computer Vision?

- Beyond Selective Tissue Sampling in Breast Cancer

Ashraf et al’s (2012) Identifying Intrinsic Cancer Imaging Phenotypes
Why study Computer Vision?

- Generate image captions automatically

A small plane parked in a field with trees in the background.

[Source: L. Zitnick, NIPS’14 Workshop on Learning Semantics]
Why study Computer Vision?

- Generate image captions automatically

A man with a colorful umbrella walking down a street.

[Source: L. Zitnick, NIPS’14 Workshop on Learning Semantics]
Why study Computer Vision?

- Generate image captions automatically

A herd of giraffes walk down the street in the middle of some trees.

[Source: L. Zitnick, NIPS’14 Workshop on Learning Semantics]
Why study Computer Vision?

- Generate image captions automatically

Why study Computer Vision?

- Have a computer do math for you

Figure: Photomath: https://photomath.net/
Why study Computer Vision?

- Fingerprint recognition

[Source: S. Lazebnik]
Why study Computer Vision?

- You can do some movie-like Forensics

Figure: Source: Nayar and Nishino, Eyes for Relighting

[Source: N. Snavely]
Why study Computer Vision?

[Source: N. Snavely]
Why study Computer Vision?

**Figure**: Source: Nayar and Nishino, Eyes for Relighting

[Source: N. Snavely]
Why study Computer Vision?

- Recognizing movie posters (in mobile phones)

iPhone Apps: kooaba (www.kooaba.com)

Source: S. Lazebnik
Why study Computer Vision?

- Games, games & games: 3D Pose Estimation with Depth Sensors

[Source: Microsoft Kinect]
How It All Began...
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".
50 years and thousands of PhDs later...

Popular benchmarks:

### Car

<table>
<thead>
<tr>
<th>Method</th>
<th>Setting</th>
<th>Code</th>
<th>Moderate</th>
<th>Easy</th>
<th>Hard</th>
<th>Runtime</th>
<th>Environment</th>
</tr>
</thead>
<tbody>
<tr>
<td>DenseBox2</td>
<td></td>
<td></td>
<td>89.32%</td>
<td>93.94%</td>
<td>79.81%</td>
<td>5s</td>
<td>GPU @ 2.5 Ghz (C/C++)</td>
</tr>
<tr>
<td>DPM</td>
<td></td>
<td></td>
<td>88.79%</td>
<td>91.31%</td>
<td>77.73%</td>
<td>x s</td>
<td>GPU @ 1.5 Ghz (Matlab + C/C++)</td>
</tr>
<tr>
<td>3DOP</td>
<td></td>
<td></td>
<td>88.64%</td>
<td>93.04%</td>
<td>79.10%</td>
<td>3s</td>
<td>GPU @ 2.5 Ghz (Matlab + C/C++)</td>
</tr>
</tbody>
</table>


### Cyclist

<table>
<thead>
<tr>
<th>Method</th>
<th>Setting</th>
<th>Code</th>
<th>Moderate</th>
<th>Easy</th>
<th>Hard</th>
<th>Runtime</th>
<th>Environment</th>
</tr>
</thead>
<tbody>
<tr>
<td>3DOP</td>
<td></td>
<td></td>
<td>68.94%</td>
<td>78.39%</td>
<td>61.37%</td>
<td>3s</td>
<td>GPU @ 2.5 Ghz (Matlab + C/C++)</td>
</tr>
<tr>
<td>Regionlets</td>
<td></td>
<td></td>
<td>58.72%</td>
<td>70.41%</td>
<td>51.83%</td>
<td>1s</td>
<td>&gt;8 cores @ 2.5 Ghz (C/C++)</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Method</th>
<th>Setting</th>
<th>Code</th>
<th>mean</th>
<th>aeroplane</th>
<th>bicycle</th>
<th>bird</th>
<th>boat</th>
<th>bottle</th>
<th>bus</th>
<th>car</th>
<th>cat</th>
<th>chair</th>
<th>cow</th>
<th>dining table</th>
<th>dog</th>
<th>horse</th>
<th>motorbike</th>
<th>person</th>
<th>potted plant</th>
<th>sheep</th>
<th>sofa</th>
<th>train</th>
<th>tv monitor</th>
<th>submission date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fast R-CNN + YOLO [?]</td>
<td></td>
<td></td>
<td>70.8</td>
<td>82.7</td>
<td>77.7</td>
<td>74.3</td>
<td>59.1</td>
<td>47.1</td>
<td>78.0</td>
<td>73.1</td>
<td>89.2</td>
<td>49.6</td>
<td>74.3</td>
<td>55.9</td>
<td>87.4</td>
<td>79.8</td>
<td>82.2</td>
<td>75.3</td>
<td>43.1</td>
<td>51.4</td>
<td>67.8</td>
<td>81.9</td>
<td>65.6</td>
<td>05-Jun-2015</td>
</tr>
<tr>
<td>Fast R-CNN VGG16 extra data [?]</td>
<td></td>
<td></td>
<td>68.8</td>
<td>82.0</td>
<td>77.8</td>
<td>71.6</td>
<td>55.3</td>
<td>42.4</td>
<td>77.3</td>
<td>71.7</td>
<td>89.3</td>
<td>44.5</td>
<td>72.1</td>
<td>53.7</td>
<td>87.7</td>
<td>80.0</td>
<td>82.5</td>
<td>72.7</td>
<td>36.6</td>
<td>68.7</td>
<td>65.4</td>
<td>81.1</td>
<td>62.7</td>
<td>18-Apr-2015</td>
</tr>
<tr>
<td>segDeepM [?]</td>
<td></td>
<td></td>
<td>67.2</td>
<td>82.3</td>
<td>75.2</td>
<td>67.1</td>
<td>50.7</td>
<td>49.8</td>
<td>71.1</td>
<td>69.6</td>
<td>88.2</td>
<td>42.5</td>
<td>71.2</td>
<td>50.0</td>
<td>85.7</td>
<td>76.6</td>
<td>81.8</td>
<td>69.3</td>
<td>41.5</td>
<td>51.9</td>
<td>62.2</td>
<td>73.2</td>
<td>64.6</td>
<td>29-Jan-2015</td>
</tr>
<tr>
<td>BabyLearning [?]</td>
<td></td>
<td></td>
<td>63.8</td>
<td>77.7</td>
<td>73.8</td>
<td>62.3</td>
<td>48.8</td>
<td>45.4</td>
<td>67.3</td>
<td>67.0</td>
<td>80.3</td>
<td>41.3</td>
<td>70.8</td>
<td>49.7</td>
<td>79.5</td>
<td>74.7</td>
<td>78.6</td>
<td>64.5</td>
<td>36.0</td>
<td>69.9</td>
<td>55.7</td>
<td>70.4</td>
<td>61.7</td>
<td>12-Nov-2014</td>
</tr>
</tbody>
</table>

50 years and thousands of PhDs later...

- Algorithms work pretty well
- Still some embarrassing mistakes...
- The general vision problem is not yet solved

Where pink means “person”

[This pic is from 2014]
Why is vision hard?
Why is vision hard?

- Half of the cerebral cortex in primates is devoted to processing visual information. This is a lot. Means that vision has to be pretty hard!
Why is vision hard?

All this is dog... [slide adopted from: R. Urtasun]
Why is vision hard?

Biederman, 1987

~10,000 to 30,000

[slide credit: R. Urtasun]
Why is vision hard?

Lots of data to process:

- Thousands to millions of pixels in an image
- 100 hours of video added to YouTube per minute [source: YouTube]
- Over 6 billion hours of video are watched each month on YouTube – almost an hour for every person on Earth [source: YouTube]
Why is vision hard?

Lots of data to process:

- \(~ 5000\) new tagged photos added to Flickr per minute (7M per day)
- \(~ 60M\) photos uploaded to Instagram every day [source: Instagram]

How many photos are uploaded to Flickr every day, month, year?

<table>
<thead>
<tr>
<th>Year</th>
<th>Photos Uploaded per Month</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004</td>
<td>10,000</td>
</tr>
<tr>
<td>2005</td>
<td>15,000</td>
</tr>
<tr>
<td>2006</td>
<td>20,000</td>
</tr>
<tr>
<td>2007</td>
<td>25,000</td>
</tr>
<tr>
<td>2008</td>
<td>30,000</td>
</tr>
<tr>
<td>2009</td>
<td>35,000</td>
</tr>
<tr>
<td>2010</td>
<td>40,000</td>
</tr>
<tr>
<td>2011</td>
<td>45,000</td>
</tr>
<tr>
<td>2012</td>
<td>50,000</td>
</tr>
<tr>
<td>2013</td>
<td>55,000</td>
</tr>
<tr>
<td>2014</td>
<td>60,000</td>
</tr>
</tbody>
</table>

Instagram:

- Monthly Actives: 200M
- People Outside U.S.: 65%+
- Photos Shared: 20B
- Likes Daily: 1.6B
- Average Photos Per Day: 60M
Why is vision hard?

- Human vision seems to work quite well.
- How well does it really work?
- Let’s play some games!
How good are humans?

- Which square is lighter, A or B?

[Slide credit: A. Torralba]
How good are humans?

Which square is lighter, A or B?

[Slide credit: A. Torralba]
How good are humans?

Figure: 2006 Walt Anthony

- Which red line is longer?

[Slide credit: A. Torralba]
How good are humans?

Figure: 2006 Walt Anthony

- Which red line is longer?

[Slide credit: A. Torralba]
How good are humans?

Figure: Chabris & Simons

- Count the number of times the white team pass the ball
- Concentrate, it’s difficult!
How good are humans?

Figure: Torralba et al.

- Can you describe what’s going on in the video?
How good are humans?

Figure: Torralba et al.

- Can you describe what’s going on in the video?
What do I need to become a good Computer Vision researcher?

- Some math knowledge
- Good programming skills
- Imagination
- Even better intuition
- Lots of persistence
- Some luck always helps