CSC420: Intro to Image Understanding

Introduction

Sanja Fidler

January 3, 2017
The Team

- **Instructor:**
  Sanja Fidler (fidler@cs.toronto.edu)

- **Office:** DH 3094

- **Office hours:** Monday 1-2pm, or by appointment

- **TAs:**
  Shenlong Wang (slwang@cs.toronto.edu)
  Hang Chu (chuhang1122@gmail.com)
Course Information

- **Class time**: Monday at 2-4pm
- **Location**: DH 2010
- **Tutorials**: Monday at 4-5pm in DH 2010, demos and Q&A, we’ll do it on demand
- **Class Website**: 

- The class will use Piazza for **announcements** and **discussions**: 
  https://piazza.com/utoronto.ca/winter2017/csc420

- 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

- **Class time**: Monday at 2-4pm
- **Location**: DH 2010
- **Tutorials**: Monday at 4-5pm in DH 2010, demos and Q&A, we’ll do it on demand
- **Class Website**:
  

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

- 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
● **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

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

- Each student expected to complete 4 assignments and a project

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

- Each student expected to complete 4 assignments and a project

**Assignments:**

- Short *theoretical questions* and *programming exercises*
- Will be given roughly every *two weeks* (starting 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 *individually* or in *pairs*
Requirements

- Each student expected to complete 4 assignments and a project

**Assignments:**
- Short *theoretical questions* and *programming exercises*
- Will be given roughly every *two weeks* (starting 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 *individually* or in *pairs*
Grading

- **Grade breakdown**
  - **Assignments**: 60% (15% each)
  - **Project**: 40%

- For the project you will need to hand in a:
  - Short project proposal
  - Project report
  - Project presentation (oral)

- I will be asking questions about relevant part of the material during project presentations
### Term Work Dates

<table>
<thead>
<tr>
<th>Term Work</th>
<th>Post Date</th>
<th>Due Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assignment 1</td>
<td>Jan 10</td>
<td>Jan 17</td>
</tr>
<tr>
<td>Assignment 2</td>
<td>Jan 24</td>
<td>Jan 31</td>
</tr>
<tr>
<td>Assignment 3</td>
<td>Feb 7</td>
<td>Feb 14</td>
</tr>
<tr>
<td>Assignment 4</td>
<td>Feb 28</td>
<td>March 7</td>
</tr>
<tr>
<td>Project Report</td>
<td></td>
<td>First week of April</td>
</tr>
<tr>
<td>Project Presentation</td>
<td></td>
<td>First week of April</td>
</tr>
</tbody>
</table>

- All dates are for 2017. ;)
- Dates are approximate
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 and Python
- 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 the 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>Jan 2</td>
<td>Intro</td>
</tr>
<tr>
<td>2</td>
<td>Jan 9</td>
<td>Linear filters, edges</td>
</tr>
<tr>
<td>3</td>
<td>Jan 16</td>
<td>Image features</td>
</tr>
<tr>
<td>4</td>
<td>Jan 23</td>
<td>Keypoint detection</td>
</tr>
<tr>
<td>5</td>
<td>Jan 30</td>
<td>Matching</td>
</tr>
<tr>
<td>6</td>
<td>Feb 6</td>
<td>Segmentation</td>
</tr>
<tr>
<td>7</td>
<td>Feb 13</td>
<td>Grouping</td>
</tr>
<tr>
<td>8</td>
<td>Feb 20</td>
<td>Object recognition</td>
</tr>
<tr>
<td>9</td>
<td>Feb 27</td>
<td>Object detection</td>
</tr>
<tr>
<td>10</td>
<td>March 6</td>
<td>Object detection, Neural Networks</td>
</tr>
<tr>
<td>11</td>
<td>March 13</td>
<td>Stereo, multi-view</td>
</tr>
<tr>
<td>12</td>
<td>March 20</td>
<td>Stereo, multi-view</td>
</tr>
<tr>
<td>12</td>
<td>March 27</td>
<td>Recognition in 3D</td>
</tr>
<tr>
<td>13</td>
<td>April ?</td>
<td>Project Presentations</td>
</tr>
</tbody>
</table>
Introduction
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

[tex 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

[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/material properties?

[Image of a room with a bed, shelves, and colorful objects]
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/material 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/material properties?
- What actions are taking place?
“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**

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
“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
“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” Image Understanding?

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

  ![Image of a kitchen](image1.png)

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

  ![Image of a room](image2.png)

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

  ![Image of a room](image3.png)

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

  ![Image of a room](image4.png)

  **Q:** How many lights are on?
  **A:** 6
Full understanding of an image? **You can answer any question about it**
Why study Computer Vision?
Why study Computer Vision?

- Because it is challenging and fun

Jialiang Wang’s (4th undergraduate year, UofT) video about his summer research in computer vision

Why study Computer Vision?

- You are curious how to one day make the robot walk your dog

http://www.cs.toronto.edu/~fidler/videos/robotsmovies.mov
Why study Computer Vision?

- ... and fold your laundry

https://www.youtube.com/watch?v=gy5g33S0Gzo

https://www.youtube.com/watch?v=KKUaVzf3Oqw
Why study Computer Vision?

- ... and drive you to work

Amnon Shashua’s Mobileye autonomous driving system

https://www.youtube.com/watch?v=4fxFDypHZLs
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

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

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

http://www.cs.cmu.edu/~om3d/

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

- Fancy visualization and game analysis in sports
Why study Computer Vision?

- Fancy visualization and special effects in movies

Why study Computer Vision?

- Reconstruct the world in 3D from online photos!

https://www.youtube.com/watch?v=IgBQCoEfiMs

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?

- How Fashionable Are You?

**Figure:** An example of a post on http://www.chictopia.com. We crawled the site for 180K posts.
Why study Computer Vision?

- How Fashionable Can You Become?

**Figure:** Examples of recommendations provided by our model. The parenthesis we show the fashionability scores.

[E. Simo-Serra, S. Fidler, F. Moreno, R. Urtasun. CVPR’15.]
Why study Computer Vision?


The lady's upper-clothes contain the pattern of flowers
Why study Computer Vision?


The woman is wearing a blue short-sleeved T-shirt and blue jeans
Why study Computer Vision?


A woman wearing a black overcoat and white shorts
Why study Computer Vision?

Why study Computer Vision?

- Crazy media attention!!!
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?

- 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 others worse)

[Khosla, Bainbridge, Oliva, Torralba, Modifying the Memorability of Face Photographs, ICCV 2013]
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

[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/](https://photomath.net/), [http://www.youtube.com/watch?v=X1bVB50mIh4](http://www.youtube.com/watch?v=X1bVB50mIh4)
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?

- Some more CSI

![Image with captions: a) Input (occluder present) b) Reference (occluder absent)]

- Can you see something on the wall?

Torralba & Freeman, CVPR’12
Why study Computer Vision?

- Some more CSI

---

a) Input (occluder present)  
b) Reference (occluder absent)

c) Difference image (b-a)  
d) Crop upside down  
e) True view
Why study Computer Vision?

- Object recognition (in mobile phones)

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

- There is opportunity for fame & glory, and of course serious $$$

Andrew Ng  
Stanford Univ.  

Yann Lecun  
NYU  

Rob Fergus  
NYU  

Noah Snavelly  
Cornell Univ.
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>
<th>Compare</th>
</tr>
</thead>
<tbody>
<tr>
<td>DuExe</td>
<td></td>
<td></td>
<td>92.65%</td>
<td>91.43%</td>
<td>86.18%</td>
<td>4 s</td>
<td>GPU @ 2.5 GHz (C/C++)</td>
<td></td>
</tr>
<tr>
<td>RV-CNN</td>
<td></td>
<td></td>
<td>91.67%</td>
<td>91.28%</td>
<td>85.43%</td>
<td>3.5 s</td>
<td>GPU @ 2.5 GHz (Python + C/C++)</td>
<td></td>
</tr>
<tr>
<td>e2e</td>
<td></td>
<td></td>
<td>91.28%</td>
<td>91.06%</td>
<td>85.66%</td>
<td>4 s</td>
<td>GPU @ 2.5 GHz (C/C++)</td>
<td></td>
</tr>
<tr>
<td>Genome</td>
<td></td>
<td></td>
<td>90.63%</td>
<td>90.85%</td>
<td>85.82%</td>
<td></td>
<td>GPU @ 2.5 GHz (C/C++)</td>
<td></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>
<th>Compare</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pie</td>
<td></td>
<td></td>
<td>76.25%</td>
<td>84.62%</td>
<td>67.57%</td>
<td>1.2 s</td>
<td>1 core @ 2.5 GHz (C/C++)</td>
<td></td>
</tr>
<tr>
<td>SAIL</td>
<td></td>
<td></td>
<td>76.13%</td>
<td>83.88%</td>
<td>66.60%</td>
<td>0.15 s</td>
<td>GPU @ &gt;3.3 GHz (Python + C/C++)</td>
<td></td>
</tr>
<tr>
<td>TiCNN</td>
<td></td>
<td></td>
<td>75.83%</td>
<td>84.28%</td>
<td>66.50%</td>
<td>0.5 s</td>
<td>GPU @ 2.5 GHz (Matlab + C/C++)</td>
<td></td>
</tr>
<tr>
<td>TuSimple</td>
<td></td>
<td></td>
<td>75.59%</td>
<td>84.15%</td>
<td>66.35%</td>
<td>1.6 s</td>
<td>GPU @ 2.5 GHz (C/C++)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Class</th>
<th>Mean</th>
<th>Aero</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>Motor Bike</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 [7]</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>71.4</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 [7]</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>64.0</td>
<td>18-Apr-2015</td>
</tr>
<tr>
<td>segDeepM [7]</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>71.9</td>
<td>62.2</td>
<td>73.2</td>
<td>64.6</td>
<td>29-Jan-2015</td>
</tr>
<tr>
<td>BabyLearning [7]</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>64.1</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...
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:

- \( \sim 5000 \) new tagged photos added to Flickr per minute (7M per day)
- \( \sim 60M \) photos uploaded to Instagram every day [source: Instagram]
Exploit so Much Data!

Figure: Vemodalen: The Fear That Everything Has Already Been Done, https://www.youtube.com/watch?v=8ftDjebw8aA

[Source: L. Zitnick, NIPS’14 Workshop on Learning Semantics]
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:*** Ames room

- Assumptions can be wrong

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

**Figure:** Chabris & Simons, https://www.youtube.com/watch?v=vJG698U2Mvo

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

Figure: Simons et al., http://www.perceptionweb.com/perception/perc1000/a_d_ex1.mov (more videos here: http://www.perceptionweb.com/misc.cgi?id=p3104)

- Is something happening in the picture?
How good are humans?

Figure: Torralba et al., http://people.csail.mit.edu/torralba/courses/6.870/slides/blur.avi

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

Figure: Torralba et al., http://people.csail.mit.edu/torralba/courses/6.870/slides/highres.avi

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

- Technical capabilities
- Good programming skills
- Imagination
- Even better intuition
- Lots of persistence
- Some luck always helps