Lecture 1: Course Overview; Images
Norms
Goals Slides

• A slide like this will appear at the beginning of each lecture.

• This is my attempt to be transparent about what I want you to get out of the lecture.

• This can be helpful for you (when studying for exams) as well as me (when writing exams).

• When I introduce key terminology, I (try to) highlight it in blue like this.
Goals

- Understand how vision tasks live on a spectrum from "low level" to "high level" vision.

- Appreciate why computer vision is hard.

- Understand how images are represented:
  - On a computer
  - In math

- Be able to write down mathematical image transformations that perform simple photometric or geometric manipulations of images functions.

- Know what is meant by image noise.
Course Overview

What does this course cover?
Visual Perception

What can you tell me about this image? or,
What might you want to know about this image?

send suggestions to #general
Levels of Vision

What can you tell me about this image? or,
What might you want to know about this image?

What would a noise-free image look like?
Where are the edges?
Where are there straight lines?
Which patches of this image are distinctive?
How would the unblurred background appear?
Is this the same scene as another image?
How far from the camera is each point?
What is the 3D shape of the subject?
Is the image level? Which way is up?
Which groups of pixels "belong" together?
Which pixels belong to the subject?
What is the subject?
What breed of dog is the subject?
What is the subject's emotional state?
Is the subject a Very Good Boi?
Levels of Vision

What can you tell me about this image? or,
What might you want to know about this image?

What would a noise-free image look like?
Where are the edges?
Where are there straight lines?
Which patches of this image are distinctive?
How would the unblurred background appear?
Is this the same scene as another image?
How far from the camera is each point?
What is the 3D shape of the subject?
Is the image level? Which way is up?
Which groups of pixels "belong" together?
Which pixels belong to the subject?
What is the subject?
What breed of dog is the subject?
What is the subject's emotional state?
Is the subject a Very Good Boi?
# Levels of Vision

**Subfield/topic**

- Filtering / Image Processing
  - Feature detection
- Computational Photography
  - Image matching / stitching
- Geometric vision
- Segmentation
- Semantic understanding

**Example(s)**

**"low-level"**

- What would a noise-free image look like?
- Where are the edges?
- Where are there straight lines?
- Which patches of this image are distinctive?
- How would the unblurred background appear?
- Is this the same scene as another image?
- How far from the camera is each point?
- What is the 3D shape of the subject?
- Is the image level? Which way is up?
- Which groups of pixels "belong" together?
- Which pixels belong to the subject?
- What is the subject?
- What breed of dog is the subject?
- What is the dog's emotional state?

**"high-level"**
In this course...

What would a noise-free image look like?
Where are the edges?
Where are there straight lines?
Which patches of this image are distinctive?
How would the unblurred background appear?
Is this the same scene as another image?
How far from the camera is each point?
What is the 3D shape of the subject?
Is the image level? Which way is up?
Which groups of pixels "belong" together?
Which pixels belong to the subject?
What is the subject?
What breed of dog is the subject?
What is the dog's emotional state?
Course Overview

5ish major topics:

• Image filtering
• Feature detection
• Panorama stitching
• Depth estimation via stereo vision
• Image recognition via deep learning
Course Overview

5ish major topics:

- Image filtering
- Feature detection
- Panorama stitching
- Depth estimation via stereo vision
- Image recognition via deep learning

"low-level"

"high-level"
Course Overview

5ish major topics:

- Image filtering
- Feature detection
- Panorama stitching
- Depth estimation via stereo vision
- Image recognition via deep learning

"low-level"

"high-level"
Course Overview

5ish major topics:

- Image filtering
- Feature detection
- Panorama stitching
- Depth estimation via stereo vision
- Image recognition via deep learning
Course Overview

5ish major topics:

- Image filtering
- Feature detection
- Panorama stitching
- Depth estimation via stereo vision
- Image recognition via deep learning
Vision is hard?

What would a noise-free image look like?
Where are the edges?
Where are there straight lines?
Which patches of this image are distinctive?
How would the unblurred background appear?
Is this the same scene as another image?
How far from the camera is each point?
What is the 3D shape of the subject?
Is the image level? Which way is up?
Which groups of pixels "belong" together?
Which pixels belong to the subject?
What is the subject?
What breed of dog is the subject?
What is the dog's emotional state?
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.

In CS, it can be hard to explain the difference between the easy and the virtually impossible.
Introducing: Flickr PARK or BIRD

Zion National Park Utah by Les Haines (CC) BY

Secretary Bird by Bill Gracey (CC) BY-NC-ND

convolution + nonlinearity
max pooling

convolution + pooling layers

vec

fully connected layers
Nx binary classification

bird
P_{bird}
sunset
P_{sunset}
dog
P_{dog}
cat
P_{cat}

...
Why is vision hard?
Why is vision hard?

Viewpoint variation
Why is vision hard?

Viewpoint variation

Scale
Why is vision hard?

Viewpoint variation

Illumination

Scale
Why is vision hard?
Why is vision hard?

Intra-class variation
Why is vision hard?

Intra-class variation

Motion (Source: S. Lazebnik)
Why is vision hard?

- Intra-class variation
- Motion (Source: S. Lazebnik)
- Occlusion
Why is vision hard?

- Intra-class variation
- Motion (Source: S. Lazebnik)
- Background clutter
- Occlusion
Why is vision hard?

The state of Computer Vision and AI: we are really, really far.

Oct 22, 2012

The picture above is funny.
But for me it is also one of those examples that make me sad about the outlook for AI and for Computer Vision. What would it take for a computer to understand this image as you or I do? I challenge you to think explicitly of all the pieces of knowledge that have to fall in place for it to make sense. Here is my short attempt:

- You recognize it is an image of a bunch of people and you understand they are in a hallway.
- You recognize that there are 3 mirrors in the scene so some of those people are "fake" replicas from different viewpoints.
- You recognize Obama from the few pixels that make up his face. It helps that he is in his suit and that he is surrounded by other people with suits.
- You recognize that there's a person standing on a scale, even though the scale occupies only very few white pixels that blend with the background. But, you've used the person's pose and knowledge of how people interact with objects to figure it out.
- You recognize that Obama has his foot positioned just slightly on top of the scale. Notice the language I'm using: It is in terms of the 3D structure of the scene, not the position of the leg in the 2D coordinate system of the image.
- You know how physics works: Obama is leaning in on the scale, which applies a force on it. Scale measures force that is applied on it, that's how it works => it will over-estimate the weight of the person standing on it.
- The person measuring his weight is not aware of Obama doing this. You derive this because you know his pose, you understand that the field of view of a person is finite, and you understand that he is not very likely to sense the slight push of Obama's foot.
- You understand that people are self-conscious about their weight. You also understand that he is reading off the scale measurement, and that shortly the over-estimated weight will confuse him because it will probably be much higher than what he expects. In other words, you reason about implications of the events that are about to unfold seconds after this photo was taken, and especially about the thoughts and how they will develop inside people's heads. You also reason about what pieces of information are available to people.
- There are people in the back who find the person's imminent confusion funny. In other words you are reasoning about state of mind of people, and their view of the state of mind of another person. That's getting frighteningly meta.
- Finally, the fact that the perpetrator here is the president makes it maybe even a little more funnier. You understand what actions are more or less likely to be undertaken by different people based on their status and identity.
What is an image?

Computationally speaking...

A grayscale image is a 2D array of numbers.
What is an image?

Computationally speaking...

A grayscale image is a 2D array of numbers.
What is an image?

Computationally speaking...

A **grayscale** image is a 2D array of numbers.

usually one byte per pixel
What is an image?

Mathematically speaking...

A grayscale image is a function $f$, from $\mathbb{R}^2 \rightarrow \mathbb{R}$

$$f(x, y)$$
What's the difference?

The computational representation is a **sampled** version of the (ideal) mathematical representation.

(ideally) continuous  
step function
What's the difference?

The computational representation is a **sampled** version of the (ideal) mathematical representation.

(ideally) continuous  
step function

(we can also still write a step function that represents the sampled version)
How do images happen?

more on this process later...
What is a color image?

3 grayscale images, each representing one color channel

- Often red, green and blue (RGB)
What is a color image?

3 grayscale images, each representing one color channel
Often red, green and blue (RGB)

Computationally: HxWx3 array
What is a color image?

3 grayscale images, each representing one color channel
Often red, green and blue (RGB)

Computationally: HxWx3 array
Mathematically: \( f : \mathbb{R}^2 \rightarrow \mathbb{R}^3 \)
What is a color image?

3 grayscale images, each representing one color channel

Often red, green and blue (RGB)

Computationally: HxWx3 array

Mathematically: \( f : \mathbb{R}^2 \rightarrow \mathbb{R}^3 \)
What is a color image?

3 grayscale images, each representing one color channel

Often red, green and blue (RGB)

Computationally: HxWx3 array

Mathematically: \( f : \mathbb{R}^2 \rightarrow \mathbb{R}^3 \)

position      color
What is a color image?

3 grayscale images, each representing one color channel
Often red, green and blue (RGB)

Computationally: HxWx3 array
Mathematically: $f : \mathbb{R}^2 \to \mathbb{R}^3$

Not always RGB (HSL, Lab, ...)
Not always only 3 channels
Images are big

Efficiency matters - and not just big-O:

\( n \) is large, so even \( O(n) \) can be slow

This is a pretty small image.
It's 227 x 336 pixels, grayscale.
That's 95,842 bytes, or 93 kb
Images are big

Efficiency matters - and not just big-O:

\( n \) is large, so even \( O(n) \) can be slow

This is a pretty small image.
It's 227 x 336 pixels, grayscale.
That's 95,842 bytes, or 93 kb

A grayscale 1080p frame (1920x1080) is about 2MB.
Images are big

Efficiency matters - and not just big-O:

\( n \) is large, so even \( O(n) \) can be slow

This is a pretty small image.
It's 227 x 336 pixels, grayscale.
That's 95,842 bytes, or 93 kb

A grayscale 1080p frame (1920x1080) is about 2MB. Want color? 6MB.
Images are big

Efficiency matters - and not just big-O:

\( n \) is large, so even \( O(n) \) can be slow

This is a pretty small image.
It's 227 x 336 pixels, grayscale.
That's 95,842 bytes, or 93 kb

A grayscale 1080p frame (1920x1080) is about 2MB.
Want color? 6MB.
Want video? ~30 frames per second; 900MB for a 5s clip
Images are big

Efficiency matters - and not just big-O:

n is large, so even O(n) can be slow

This is a pretty small image.
It's 227 x 336 pixels, grayscale.
That's 95,842 bytes, or 93 kb

A grayscale 1080p frame (1920x1080) is about 2MB.
Want color? 6MB.
Want video? ~30 frames per second; 900MB for a 5s clip
Want a movie? 1296 gigabytes for a 2-hour movie.
Image Processing: What can we do to an image?

Written as a function, we can *transform* the image function to create altered functions (images):

\[ g(x, y) = f(x, y) + 20 \]

**photometric transformations**

\[ g(x, y) = f(-x, y) \]

**geometric transformations**
Image Processing:
What can we do to an image?

Written as a function, we can *transform* the image function to create altered functions (images):

\[ g(x,y) = f(x,y) + 20 \]

*photometric transformations* (increase brightness)

\[ g(x,y) = f(-x,y) \]

*geometric transformations*
Image Processing:
What can we do to an image?

Written as a function, we can *transform* the image function to create altered functions (images):

- **photometric transformations**
  
  \[ g(x, y) = f(x, y) + 20 \]  
  
  (increase brightness)

- **geometric transformations**

  \[ g(x, y) = f(-x, y) \]  
  
  (flip horizontally)
Lecture 1 Problems

• In your breakout groups:
  • Find the "Problems" linked from today's lecture on the course webpage.
  • Open the Google Doc pinned in your group's text channel
  • Collaboratively write up answers to the problems in the google doc.
Real images aren't perfect

Real images are not only sampled, but they often have noise: unwanted variations in measured intensity value.

Causes of noise (incomplete list):
- electronic variations in sensor chip
- analog-to-digital quantization
- film grain
- cosmic rays

$f(x, y)$
Real images aren't perfect

Real images are not only sampled, but they often have **noise**: unwanted variations in measured intensity value.

Causes of noise (incomplete list):
- electronic variations in sensor chip
- analog-to-digital quantization
- film grain
- cosmic rays

Often, we can assume that noise is *random*.

\[ f(x, y) \]
Real images aren't perfect

Real images are not only sampled, but they often have **noise**: unwanted variations in measured intensity value.

Causes of noise (incomplete list):
- electronic variations in sensor chip
- analog-to-digital quantization
- film grain
- cosmic rays

Often, we can assume that noise is *random*

(Other times, we can't but we do anyway)
Real images aren't perfect

- Real images are not only sampled, but they often contain noise.

\[ f(x, y) \]

How could we denoise \( f \)?