CSE 152A: Computer Vision
Manmohan Chandraker

Lecture 10: Recognition
Overall goals for the course

• Introduce fundamental concepts in computer vision

• Enable one or all of several such outcomes
  – Pursue higher studies in computer vision
  – Join industry to do cutting-edge work in computer vision
  – Gain appreciation of modern computer vision technologies

• Engage in discussions and interaction

• This is a great time to study computer vision!
Course details

• Class webpage:
  – http://cseweb.ucsd.edu/~mkchandraker/classes/CSE152A/Fall2022/

• TAs
  – Meng Song, Mallikarjun Swamy, Rishi Chandrasekaran, Vishal Vinod: {mes050, mswamy, r3chandr, vvinod}@ucsd.edu

• Tutors
  – Nick Chua, Navya Sharma, Ang Li: {nchua, n1sharma, a3li}@ucsd.edu

• Discussion section: M 3-3:50pm

• Office hours posted on course calendar

• Piazza: https://piazza.com/ucsd/fall2022/cse152a/
Self-Study Assignment

• Pick a technology area primarily driven by computer vision
  – Can pick one of these suggestions, or use anything else that you like

• Write a 1 page essay (single-spaced)
  – Can be longer (hopefully not too long)
  – Great if you include pictures (with citations)

• Example prompts (feel free to add to these or use your own):
  – How does computer vision overcome barriers or solve user needs in this technology?
  – Can you identify where knowledge of photometric or geometric image formation is used?
  – Can you identify where such knowledge does not suffice and machine learning is used?
  – Can you identify where photometric or geometric models are used along with learning?
  – What was possible in this area 10 years ago and how did computer vision advance it?
  – How do you anticipate technology in this chosen area will advance in the next 10 years?

• Due date: Nov 11, 2022
Mid-Term Discussions
Mid-Term Feedback

**Things I like:**

- Professor, TAs, tutors

**Things to improve:**

- Professor, TAs, tutors
Mid-Term Feedback

**Things I like:**
- Professor, TAs, tutors
- Lecture pace is good, review at beginning of lectures is helpful

**Things to improve:**
- Professor, TAs, tutors
- Lecture are too fast, need more examples, annotated slides
# Mid-Term Feedback

**Things I like:**
- Professor, TAs, tutors
- Lecture pace is good, review at beginning of lectures is helpful
- Homeworks are tough, but the outcomes are nice

**Things to improve:**
- Professor, TAs, tutors
- Lecture are too fast, annotated slides would be helpful
- Homeworks take long, content not covered when released
# Mid-Term Feedback

**Things I like:**

- Professor, TAs, tutors
- Lecture pace is good, review at beginning of lectures is helpful
- Homeworks are tough, but the outcomes are nice
- OHs and discussions are helpful, cover background material

**Things to improve:**

- Professor, TAs, tutors
- Lecture are too fast, annotated slides would be helpful
- Homeworks take long, content not covered when released
- OHs and discussions should be more detailed on homework
Mid-Term Feedback

**Things I like:**

- Professor, TAs, tutors
- Lecture pace is good, review at beginning of lectures is helpful
- Homeworks are tough, but the outcomes are nice
- OHs and discussions are helpful, cover background material
- Love the fact that I lose so much sleep over this class

**Things to improve:**

- Professor, TAs, tutors
- Lecture are too fast, annotated slides would be helpful
- Homeworks take long, content not covered when released
- OHs and discussions should be more detailed on homework
Mid-Term Feedback

Things I like:

• Professor, TAs, tutors
• Lecture pace is good, review at beginning of lectures is helpful
• Homeworks are tough, but the outcomes are nice
• OHs and discussions are helpful, cover background material
• Love the fact that I lose so much sleep over this class

Things to improve:

• Professor, TAs, tutors
• Lecture are too fast, annotated slides would be helpful
• Homeworks take long, content not covered when released
• OHs and discussions should be more detailed on homework
• Lectures induce sleep, won’t hurt to have some jokes
Recognition
Autonomous driving

Where is our car?
Structure from Motion
Visual SLAM

What is a safe path?
Behavior prediction
Path planning

Where are other agents?
Object detection
3D localization

Where are scene elements?
Semantic segmentation

CSE 152A, FA22: Manmohan Chandraker
So what does object recognition involve?
Verification: is that a bus?
Detection: are there cars?
Identification: is that a picture of Mao?
Object categorization
Scene and context categorization

- outdoor
- city
- traffic
- ...

CSE 152A, FA22: Manmohan Chandraker
Why is computer vision difficult?

- Viewpoint
- Lighting
- Occlusion
- Deformation
Large number of categories

10,000 to 30,000
~10,000 to 30,000
Challenges: intra-class variation

Can you design an algorithm to describe a chair?
Object categorization: the statistical viewpoint

\[ p(\text{zebra} \mid \text{image}) \]

or

\[ p(\text{no zebra} \mid \text{image}) \]

- Bayes rule:

\[
\frac{p(\text{zebra} \mid \text{image})}{p(\text{no zebra} \mid \text{image})} = \frac{p(\text{image} \mid \text{zebra})}{p(\text{image} \mid \text{no zebra})} \cdot \frac{p(\text{zebra})}{p(\text{no zebra})}
\]

- posterior ratio
- likelihood ratio
- prior ratio

CSE 152A, FA22: Manmohan Chandraker
Object categorization: the statistical viewpoint

\[
\frac{p(\text{zebra} \mid \text{image})}{p(\text{no zebra} \mid \text{image})} = \frac{p(\text{image} \mid \text{zebra})}{p(\text{image} \mid \text{no zebra})} \cdot \frac{p(\text{zebra})}{p(\text{no zebra})}
\]

- **Posterior ratio**
- **Likelihood ratio**
- **Prior ratio**

- **Discriminative methods model posterior**
- **Generative methods model likelihood and prior**

CSE 152A, FA22: Manmohan Chandraker
Discriminative

• Direct modeling of \[ \frac{p(zebra \mid image)}{p(no \ zebra \mid image)} \]
Generative

- Model $p(\text{image} | \text{zebra})$ and $p(\text{image} | \text{no zebra})$

| $p(\text{image} | \text{zebra})$ | $p(\text{image} | \text{no zebra})$ |
|----------------|----------------|
| Low            | High           |
| High           | Low            |

CSE 152A, FA22: Manmohan Chandraker
Three main issues

• Representation
  – How to represent an object category

• Learning
  – How to form the classifier, given training data

• Inference
  – How the classifier is to be used on novel data
Representation

– Generative or discriminative or hybrid
Representation

- Generative or discriminative or hybrid
- Appearance only or location and appearance
Representation

– Generative or discriminative or hybrid
– Appearance only or location and appearance
– Invariances
  • View point
  • Illumination
  • Occlusion
  • Scale
  • Deformation
  • Clutter
  • etc.
Representation

– Generative or discriminative or hybrid
– Appearance only or location and appearance
– Invariances
– Use set of features or each pixel in image
Learning

- Unclear how to model categories, so we learn what distinguishes them rather than manually specify the difference
  - Thus, the current interest in machine learning
Learning

– Unclear how to model categories, so we learn what distinguishes them rather than manually specify the difference
  • Thus, the current interest in machine learning

– Method of training: generative or discriminative
  • What are you maximizing? Likelihood (generative) or performances on train and validation set (discriminative)
Learning

- Unclear how to model categories, so we learn what distinguishes them rather than manually specify the difference
  - Thus, the current interest in machine learning
- Method of training: generative or discriminative
  - What are you maximizing? Likelihood (generative) or performances on train and validation set (discriminative)
- Level of supervision
  - Image label, bounding box, object boundaries, semantic parts

Contains a motorbike
Learning

- Unclear how to model categories, so we learn what distinguishes them rather than manually specify the difference
  - Thus, the current interest in machine learning
- Method of training: generative or discriminative
  - What are you maximizing? Likelihood (generative) or performances on train and validation set (discriminative)
- Level of supervision
  - Image label, bounding box, object boundaries, semantic parts
- Training images:
  - Issue of overfitting
Inference

– Scale or orientation range to search
– Speed
Instance Recognition
Recognizing or retrieving specific objects

Example 1: Place recognition for self-driving or robot navigation
Recognizing or retrieving specific objects

Example 2: Search photos for particular places

Find these landmarks...in these images and 1M more
Why is it difficult?

Want to find the object despite possibly large changes in scale, viewpoint, lighting and partial occlusion.
Object → Bag of ‘words’
Origin: Bag-of-words models

- Orderless document representation: frequencies of words from a dictionary  
  Salton & McGill (1983)

US Presidential Speeches Tag Cloud
http://chir.ag/phernalia/preztags/

CSE 152A, FA22: Manmohan Chandraker
Origin: Bag-of-words models


US Presidential Speeches Tag Cloud
http://chir.ag/phernalia/preztags/

CSE 152A, FA22: Manmohan Chandraker
Origin: Bag-of-words models

Bags of features for object recognition

- Works quite well for image-level classification and for recognizing object *instances*

CSE 152A, FA22: Manmohan Chandraker
Recall: matching local features

To generate **candidate matches**, find patches that have the most similar appearance (e.g., lowest SSD)

Simplest approach: compare them all, take the closest (or closest k, or within a thresholded distance)
Indexing local features

Training dataset

Test image
Bag of features: outline

1. Extract features
Bag of features: outline

1. Extract features
2. Learn “visual vocabulary”
Bag of features: outline

1. Extract features
2. Learn “visual vocabulary”
3. Quantize features using visual vocabulary
Bag of features: outline

1. Extract features
2. Learn “visual vocabulary”
3. Quantize features using visual vocabulary
4. Represent images by frequencies of “visual words”
Histogram

- A statistical tool to represent properties of a data distribution

- You own a shop, want to determine whether to hand out candies to visitors

- Collect ages of visitors:
  - [9, 34, 47, 23, 15, 18, 57, 73, 89, 20, 14, 47, 32, 75, 5, 52, 25, 32, 12, 33]
Histogram

- A statistical tool to represent properties of a data distribution

- You own a shop, want to determine whether to hand out candies to visitors
- Collect ages of visitors:
  - [9, 34, 47, 23, 15, 18, 57, 73, 89, 20, 14, 47, 93, 27, 32, 75, 5, 52, 32, 12, 33, 64, 29]

<table>
<thead>
<tr>
<th>Age range</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 – 20</td>
<td>7</td>
</tr>
<tr>
<td>21 – 40</td>
<td>6</td>
</tr>
<tr>
<td>41 – 60</td>
<td>4</td>
</tr>
<tr>
<td>61 – 80</td>
<td>5</td>
</tr>
<tr>
<td>81 – 100</td>
<td>2</td>
</tr>
</tbody>
</table>
Histogram

• A statistical tool to represent properties of a data distribution

• You own a shop, want to determine whether to hand out candies to visitors

• Collect ages of visitors:
  – [9, 34, 47, 23, 15, 18, 57, 73, 89, 20, 14, 47, 67, 93, 27, 32, 75, 5, 52, 32, 12, 33, 64, 29]

<table>
<thead>
<tr>
<th>Age range</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 – 20</td>
<td>7</td>
</tr>
<tr>
<td>21 – 40</td>
<td>6</td>
</tr>
<tr>
<td>41 – 60</td>
<td>4</td>
</tr>
<tr>
<td>61 – 80</td>
<td>5</td>
</tr>
<tr>
<td>81 – 100</td>
<td>2</td>
</tr>
</tbody>
</table>
Histogram

• A statistical tool to represent properties of a data distribution

• You own a shop, want to determine whether to hand out candies to visitors

• Collect ages of visitors:
  – [9, 34, 47, 23, 15, 18, 57, 73, 89, 20, 14, 47, 67, 93, 27, 32, 75, 5, 52, 32, 12, 33, 64, 29]

<table>
<thead>
<tr>
<th>Age range</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 – 20</td>
<td>7</td>
</tr>
<tr>
<td>21 – 40</td>
<td>6</td>
</tr>
<tr>
<td>41 – 60</td>
<td>4</td>
</tr>
<tr>
<td>61 – 80</td>
<td>5</td>
</tr>
<tr>
<td>81 – 100</td>
<td>2</td>
</tr>
</tbody>
</table>

• Enough visitors under 20, so give candies (anyway, you should always give candies)