GE 461
Introduction to Data Science
Spring 2022

Applications: Computer Vision

Hamdi Dibeklioğlu
Computer Vision

According to Hollywood
Computer Vision

Scene -> Sensor -> Interpreter -> Interpretation

Mountain
Lake
Trees
Sky
Human Vision

Scene → Sensor → Interpreter → Interpretation

Mountain
Lake
Trees
Sky
Computer Vision
Computer Vision

*Computer vision* tries to get computers to extract information from images
Human Vision

• Can you name some things that influence what we see?
Human Vision

• Can you name some things that influence what we see?
  • What objects are where in the world (and how they are deformed)
  • Lighting conditions of the environment
  • Position and orientation of the eyes (viewpoint)
  • Your own brain!!!
Human Vision

Dorsal or “where” stream

Parietal

Spatial processing

Frontal

Occipital

Temporal

location
movement
spatial
transformations
spatial relations

Ventral or “what” stream

Object processing

color
texture
pictorial detail
shape
size
Human Vision
Human Vision

- Count the red crosses

<p>| | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>


• Count the red crosses

```
0  X  O  X  O  X  X
X  O  X  X  X  O  X
O  X  X  O  X  X  O
X  X  O  X  O  O  X
O  X  X  O  X  X  X
X  O  X  X  X  O  X
O  X  X  O  X  X  O
X  O  X  X  X  O  X
X  X  X  O  O  X  X
X  O  X  X  X  O  X
```
Human Vision
Human Vision
Human Vision

• Computer vision is much more difficult than you might initially think:
  • Your brain needs 25% of the cortex just to solve vision
  • The general computer vision problem is still largely unsolved

• Main problem: vision needs to deal with enormous variations in the signal
  • Some of these variations are relevant and others should be ignored
Gap between “pixels” and “meaning”
Computer Vision: History

• An MIT undergraduate summer project*, in 1966, aimed to solve background/foreground segmentation and object detection/classification.

• It has been 54 year and we still work on the same problems.

*Seymour A Papert. The summer vision project, MIT, 1966.
Computer Vision

- Making useful decisions about real physical objects and scenes based on images (Shapiro & Stockman, 2001)
- Extracting descriptions of the world from pictures or sequences of pictures (Forsyth & Ponce, 2003)
- Analyzing images and producing descriptions that can be used to interact with the environment (Horn, 1986)
- Designing representations and algorithms for relating images to models of the world (Ballard & Brown, 1982)
Computer Vision

• How do we describe the variations within the class “chair”?

• Invariance to some variations can be obtained using hand-crafted models

• We generally try to learn invariance to the remaining variations from examples

Slide Credit: L. van der Maaten
Object Recognition

- Observation: chairs contain relatively lots of edges
Basics: What is an image?

• Assume an image as a function, \( f \), from \( \mathbb{R}^2 \) to \( \mathbb{R} \):
  • \( f(x, y) \) gives the intensity at position \( (x, y) \)
  • Realistically, an image is defined over a rectangle:
    • \( f: [a, b] \times [c, d] \rightarrow [0,1] \)

• Color image = Three functions combined together:
  • \( f(x, y) = \begin{bmatrix} r(x, y) \\ g(x, y) \\ b(x, y) \end{bmatrix} \)
Basics: An image as a function

Bright regions are high, dark regions are low
Basics: Digital Images

- In computer vision we operate on digital (discrete) images:
  - Sample the 2D space on a regular grid
  - Quantize each sample (round to the nearest integer)
  - Each sample is a pixel (picture element)
  - If we assume each pixel as 1 byte, values range from 0 to 255
Basics: Preprocessing

• Range transformation (pixel processing):
  \[ g(x, y) = h(f(x, y)) \]

• Example?
Basics: Preprocessing

- Range transformation (pixel processing):
  \[ g(x, y) = h(f(x, y)) \]
- Example: Noise filtering

![Image of noise filtering example]
Basics: Preprocessing

- Domain transformation (geometric transform):
  \[ g(x, y) = f(h_x(x, y), h_y(x, y)) \]

- Example: ?
Basics: Preprocessing

• Domain transformation (geometric transform):
  \[ g(x, y) = f(h_x(x, y), h_y(x, y)) \]

• Example: Rotation / Translation
Basics: Analysis Pipeline

Input → Feature extraction (hand-crafted) → Learning algorithm (your favorite classifier) → Output

Car
Not a car
Surveillance

- Face Recognition
- Object Detection
- Tracking
Surveillance

- Anomaly Detection
- Action & Activity Recognition
Autonomous Vehicles

- Vehicle Collision Warning
- Pedestrian Collision Warning at day-time and night
- General Obstacle Collision Warning
- Lane Departure Warning
- Traffic Road Sign Recognition
- High-Beam Assistance
- Traffic Sign Recognition
- Red Signal Lamp Recognition
Autonomous Vehicles

- Detection & Segmentation
- Tracking
Autonomous Vehicles

- Future Prediction
Autonomous Vehicles

- Gaze Estimation
- Mood Recognition
- Fatigue Monitoring
Photography

- Face Detection / Smile Shutter
- Focus Tracking
- Image Stabilization
- Color Enhancement
Special Effects
Special Effects (3D)
Map & 3D Model Generation

- 3D reconstruction
- Structure from motion
Weather Forecasting

- Spectral image analysis
- Image segmentation
Quality Control & Monitoring
Healthcare

- Facial expression analysis
- Pose estimation
- Medical image analysis
- Motion magnification
- Subtle motion tracking
Neuromarketing

- Gaze estimation / tracking
- Facial emotion recognition
- Age & gender estimation
- Appreciation recognition
Visual Search

- https://lens.google.com/
Gaming / Human Computer Interaction

- Pose and motion tracking
- Gesture recognition
Image Analytics: Identity Recognition / Tagging

- Face detection
- Face recognition
- Kinship recognition
Image Analytics: Captioning

- A yellow plate topped with meat and broccoli.
- A zebra standing next to a zebra in a dirt field.
- A stainless steel oven in a kitchen with wood cabinets.
- Two birds sitting on top of a tree branch.
- An elephant standing next to a rock wall.
- A man riding a bike down a road next to a body of water.
Image Synthesis/Modification
Style Transfer
DEAREST GABBY,
I MISS YOU THIS MORNING. I READ
A NEW PSYCHOLOGICAL SCIENCE STUDY
ABOUT THE PROSOCIAL WELL-BEING
BENEFITS OF HANDWRITTEN THANK YOU
NOTES FOR BOTH THE EXPRESSER
AND THE RECIPIENT. APPRECIATION TURNS OUT TO
HAVE MANY BENEFITS FOR BOTH YOU AND ME.
THANK YOU FOR THIS ENGRAVED ENVELOPE. TRULY A
GIFT WITH ETERNAL THANKS OF LOVE.

August 21, 2018

CHISTOPHER KARR

THANKS
YOU
AGAIN.

KISTLER

CPU 1 V1.62
Barcode: 18000251
Serial Number: 4977889
OCR Text: SN 4977889