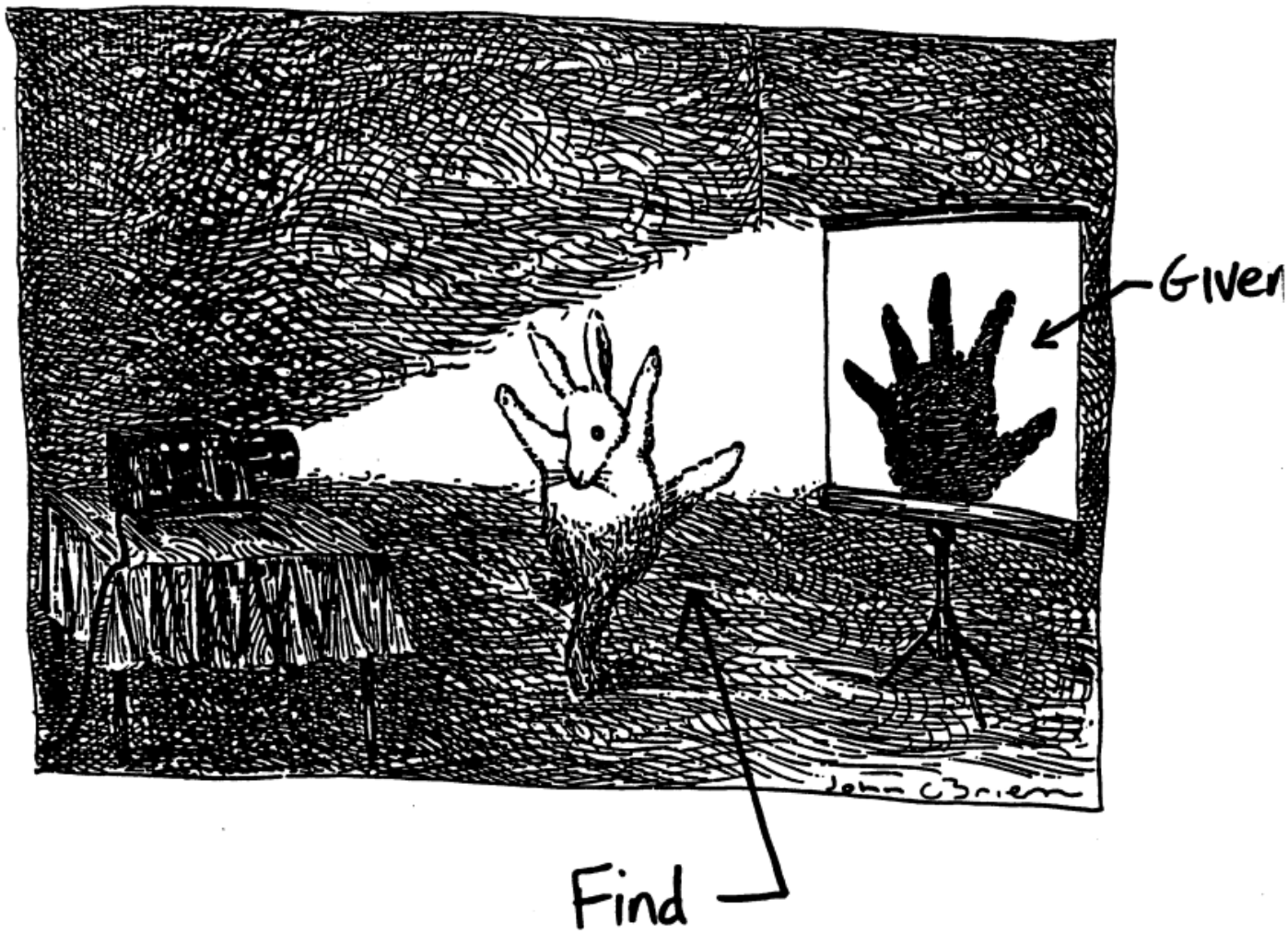
# Introduction to Computer Vision

## CSE 6367: Computer Vision

Instructor: William J. Beksi

# The Computer Vision Problem



# Why Study Computer Vision?

- Images and videos (and increasingly 3D point clouds) are everywhere

# Why Study Computer Vision?

- Images and videos (and increasingly 3D point clouds) are everywhere
- Fast-growing collection of useful applications

# Why Study Computer Vision?

- Images and videos (and increasingly 3D point clouds) are everywhere
- Fast-growing collection of useful applications
  - Building representations of the 3D world from pictures

# Why Study Computer Vision?

- Images and videos (and increasingly 3D point clouds) are everywhere
- Fast-growing collection of useful applications
  - Building representations of the 3D world from pictures
  - Automated surveillance (who's doing what)

# Why Study Computer Vision?

- Images and videos (and increasingly 3D point clouds) are everywhere
- Fast-growing collection of useful applications
  - Building representations of the 3D world from pictures
  - Automated surveillance (who's doing what)
  - Facial Recognition

# Why Study Computer Vision?

- Images and videos (and increasingly 3D point clouds) are everywhere
- Fast-growing collection of useful applications
  - Building representations of the 3D world from pictures
  - Automated surveillance (who's doing what)
  - Facial Recognition
  - Robotics (e.g. UAVs, autonomous vehicles, etc.)

# Why Study Computer Vision?

- Images and videos (and increasingly 3D point clouds) are everywhere
- Fast-growing collection of useful applications
  - Building representations of the 3D world from pictures
  - Automated surveillance (who's doing what)
  - Facial Recognition
  - Robotics (e.g. UAVs, autonomous vehicles, etc.)
- Various deep and attractive scientific mysteries, e.g. how does object recognition work?

# Why Study Computer Vision?

- Images and videos (and increasingly 3D point clouds) are everywhere
- Fast-growing collection of useful applications
  - Building representations of the 3D world from pictures
  - Automated surveillance (who's doing what)
  - Facial Recognition
  - Robotics (e.g. UAVs, autonomous vehicles, etc.)
- Various deep and attractive scientific mysteries, e.g. how does object recognition work?
- Greater understanding of human vision

# Properties of Vision

- 3D representations of the environment can be constructed

# Properties of Vision

- 3D representations of the environment can be constructed
- There are many different cues: multiple views (motion, texture, shading, etc.)

# Vision in a Single Image

- A single image is organized as a 2D grid (matrix) and contains abundant spectral information

# Vision in a Single Image

- A single image is organized as a 2D grid (matrix) and contains abundant spectral information
- Light from different parts of the spectrum is decomposed into discrete red, green, and blue (RGB) color values that we see in an image

# Color

(a) RGB

(b) R

(c) G

(d) B

# Vision in Multiple Images

- **Multiple images** allow us to obtain a richer representation of the environment

# Vision in Multiple Images

- **Multiple images** allow us to obtain a richer representation of the environment
- Where could an object appear in camera 2 (3, etc.) given it was in camera 1 (1 and 2, etc.)? (geometry of multiple views)

# Vision in Multiple Images

- **Multiple images** allow us to obtain a richer representation of the environment
- Where could an object appear in camera 2 (3, etc.) given it was in camera 1 (1 and 2, etc.)? (geometry of multiple views)
- What do we know about the world from having many eyes, or more commonly our eyes are moving? (structure from motion)

# Building Rome in a Day

# Image Processing

- **Low-level vision** operations are mainly concerned with processing the pixels of an image to extract basic features

# Image Processing

- **Low-level vision** operations are mainly concerned with processing the pixels of an image to extract basic features
- These features include corners, blobs, edges, derivatives, etc.

# Image Processing

- **Low-level vision** operations are mainly concerned with processing the pixels of an image to extract basic features
- These features include corners, blobs, edges, derivatives, etc.
- Filtering of the image is also performed at this stage, e.g. smoothing the image using a Gaussian kernel

# Edge Detection

# Finding Circles

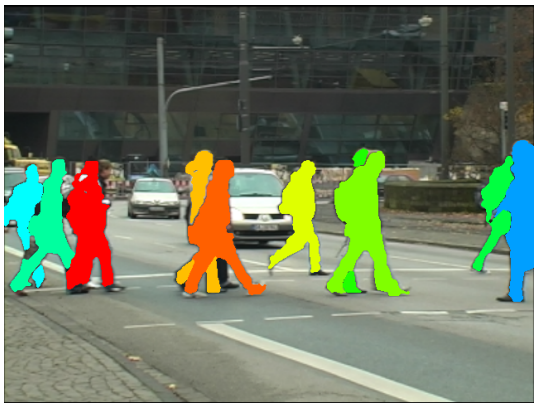
# Finding Coherent Structure

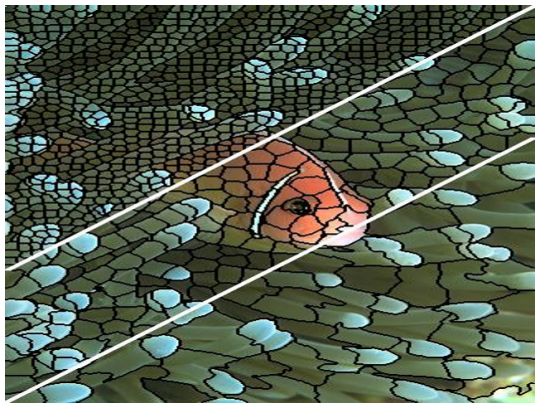
- In **mid-level vision** operations, we want to find coherent structure in order to break the image into big units

# Finding Coherent Structure

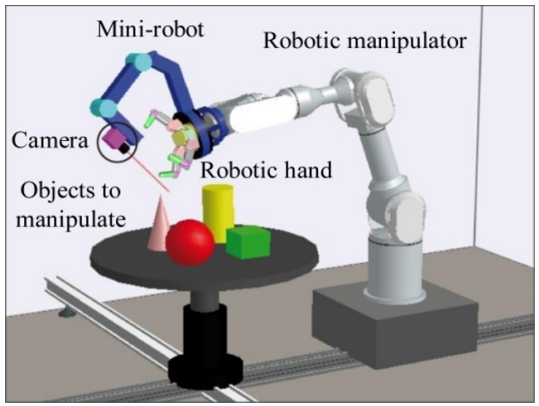
- In **mid-level vision** operations, we want to find coherent structure in order to break the image into big units
- These operations include segmentation (breaking images into useful pieces), tracking (keeping track of a moving object through a long sequence of views), etc.

# Segmentation





# Tracking


Mini-robot
Robotic manipulator
Camera
Robotic hand
Objects to manipulate

# Object and Image Geometry

- In **high-level vision** operations, we want to determine the relations between object geometry and image geometry (e.g. find the position and orientation of known objects)

# Object and Image Geometry

- In **high-level vision** operations, we want to determine the relations between object geometry and image geometry (e.g. find the position and orientation of known objects)
- We can use templates and classifiers to find objects that look the same from view to view (i.e. object recognition)

# Object and Image Geometry

- In **high-level vision** operations, we want to determine the relations between object geometry and image geometry (e.g. find the position and orientation of known objects)
- We can use templates and classifiers to find objects that look the same from view to view (i.e. object recognition)
- Furthermore, we can create hierarchical frameworks to recognize the scene in which the objects reside (i.e. scene understanding)

# Object Recognition

# Scene Understanding

# Summary

- Images provide both spectral and geometric information regarding the environment
- Computer vision can be split into low-level (image processing), mid-level (finding structure), and high-level (determining relations) operations