Artificial Intelligence

Computer Vision

Machines That See?

- Science fiction
  - HAL, Terminator, Short Circuit, I-Robot, etc.
Can a computer be designed to “see” as well as a person?

*How good is your perceptual system???

???
Past Observer Descriptions

- “bird with wings expanded”
- “an antelope/deer in a field”
- “an ice skater”
- “hood of a car”
- “impressionist painting viewed through a dirty window”
Moving Point-Lights

Which is Male, Which is Female?
Which is Male, Which is Female?

How do we “Encode” Behaviors?

HeiderSimmel.mp4
Moving Edge Fragments

Images/Videos are “Numbers”
Definition of Computer Vision

• Goal of computer vision is to make useful decisions about real physical objects and scenes based on sensed images
  – Process of discovering from images what is present in the world, where it is, and what it is doing
• Construction of scene descriptions from images
• Require representations of shape, motion, color, context, etc.

Applications

• How can computer vision be used to facilitate more intelligent systems or natural computer interfaces?
  – Recognize object, people, gestures
  – Analyze movements/activities of person
  – Identity recognition
    • From face, fingerprints, motion, etc.
  – etc.
Earth viewers (3D modeling)

Image from Microsoft's Virtual Earth
(see also: Google Earth)

Optical Character Recognition (OCR)

Technology to convert scanned docs to text
- If you have a scanner, it probably came with OCR software

Digit recognition, AT&T labs
http://www.research.att.com/~yann/

License plate readers
http://en.wikipedia.org/wiki/Automatic_number_plate_recognition
Face Detection

- Digital cameras and cell phones detect faces

Smile Detection

The Smile Shutter flow
Imagine a camera smart enough to catch every smile! In Smile Shutter Mode, your Cyber-shot® camera can automatically trip the shutter at just the right instant to catch the perfect expression.

Sony Cyber-shot® T70 Digital Still Camera
Sports

Sportvision first down line
Nice [explanation](http://www.howstuffworks.com) on www.howstuffworks.com

Sportvision Pass Track

Robotics

NASA’s Mars Spirit Rover

Medical Imaging

What’s Needed?

• Need appropriate data representations and algorithms (for given application)

• How do we capture/represent/use Color, Shape, Motion, Texture, Object Types, Depth, …

• Let’s take a look at a few approaches…
Stereo Vision

Pinhole Camera Model

- Pinhole camera model results from reducing aperture diameter to infinitesimally small point
  - Simple model and all objects in focus
- Rays connect image plane to object through pinhole
  - Object is imaged “upside-down” on image plane

We will use this virtual model
Problem With Single Views?

- Structure and depth are inherently ambiguous from single views

This is called “**Forced Perspective**”
(used in Harry Potter movies!)

Images from Luna Lazebnik

Problem With Single Views?

- Structure and depth are inherently ambiguous from single view

![Diagram](image)
Extracting 3-D Information

- **Stereo:**
  - Shape from difference between two views
  - Infer 3-D shape of scene from two (multiple) images from different viewpoints
    - Need information on camera pose (calibration) and image point correspondences

Main idea:

Depth From Stereo

- Assume parallel optical axes, known camera parameters (i.e., calibrated cameras). **What is expression for Z (depth from pinholes \(O_l, O_r\))?**

Similar triangles \((p, p, p_r)\) and \((O_l, p, O_r)\):

\[
\frac{T + x_r - x_l}{Z - f} = \frac{T}{Z}
\]

\[
Z = f \frac{T}{x_l - x_r}
\]
Depth From Disparity

image I(x,y)  Disparity map D(x,y)  image I’(x’,y’)

So if we could find the corresponding points in two images, we could estimate relative depth (methods exist to achieve this).

Finding “Interesting” Points
Application: Image Stitching

[Images of mountain scenes before and after stitching]
Application: Image Stitching

Application: Image Stitching
Features from Accelerated Segment Test (FAST)

• Compare local neighborhood around each pixel to determine if pixel is a good feature point

• For each pixel $x$:
  - Look at the pixels on the border of a circle with radius $r$ around $x$
  - Let $n$ be the largest number of contiguous pixels whose intensities are either
    
    \[
    \text{“all } n\text{”} \geq (I_x + T) \quad \text{or} \quad \text{“all } n\text{”} \leq (I_x - T)
    \]
    
    $I_x$ is the intensity at pixel $x$, and $T$ is an intensity delta/range.
  - If $n \geq n^*$ then the pixel is considered a feature
  - Original paper (Rosten 2006) suggests using $r = 3$ (yields 16 pixels on border) and $n^* = 12$
    • Later results suggest using $r = 3$ (yields 16 pixels on border) and $n^* = 9$
      provides best detector.

FAST Example

• Radius $r = 3$ and some $T$

\[
\begin{array}{cccccccccccc}
I_1 & I_2 & I_3 & I_4 & I_5 & I_6 & I_7 & I_8 & I_9 & I_{10} & I_{11} & I_{12} & I_{13} & I_{14} & I_{15} & I_{16} \\
\end{array}
\]

\[
\begin{array}{cccccccccccc}
16 & 1 & 2 & & & & & & & & & & & & \\
15 & & & 3 & & & & & & & & & & & \\
14 & & & & 4 & & & & & & & & & & \\
13 & & & & & 5 & & & & & & & & & \\
12 & & & & & & 6 & & & & & & & & \\
11 & & & & & & & 7 & & & & & & & \\
10 & & & & & & & & 8 & & & & & & \\
9 & & & & & & & & & 9 & & & & & \\
\end{array}
\]

\[
\begin{array}{cccccccccccc}
\text{A: } & I_i > (I_c + T) \\
\text{B: } & I_i < (I_c - T) \\
\text{W: otherwise} \\
\end{array}
\]
FAST Example \( (r = 3, \ n^* = 12) \)

Input

\( T = 10 \)

\( T = 30 \)

\( T = 50 \)

FAST Example \( (r = 3, \ n^* = 9) \)

Input

\( T = 10 \)

\( T = 30 \)

\( T = 50 \)
Summary

- Computer vision is the task of making machines “see”
- Need representations/algorithms for various aspects of the visual world (e.g., color, motion, depth, …)
- Many products using computer vision in consumer market
- CSE 5524 course

Some slides adapted from T. Darrell, J. Hays, O. Grau, M. Pollefeys, K. Grauman, S. Seitz, L. Sigal, D. Fleet, and A. Hertzmann