Computer Vision: CSE 803

A brief intro
Today's Class

- Imaging process
- Image processing operation
- Perceive 3D from images
- CV applications
- Homework #1 brief
Goal of Computer Vision

- Make useful decisions about real physical objects and scenes based on sensed images.
- Alternative (Aloimonos and Rosenfeld): goal is the construction of scene descriptions from images.
- How do you find the door to leave?
- How do you determine if a person is friendly or hostile? .. an elder? .. a possible mate?
Critical Issues

- **Sensing**: how do sensors obtain images of the world?
- **Information/features**: how do we obtain color, texture, shape, motion, etc.?
- **Representations**: what representations should/does a computer [or brain] use?
- **Algorithms**: what algorithms process image information and construct scene descriptions?
Images: 2D Projections of 3D

- 3D world has color, texture, surfaces, volumes, light sources, objects, motion, betweenness, adjacency, connections, etc.
- 2D image is a projection of a scene from a specific viewpoint; many 3D features are captured, some are not.
- Brightness or color = g(x,y) or f(row, column) for a certain instant of time.
- Images indicate familiar people, moving objects or animals, health of people or machines.
Image Receives Reflections

- Light reaches surfaces in 3D
- Surfaces reflect
- Sensor element receives light energy
- Intensity matters
- Angles matter
- Material matters
Simple Objects: Simple Image?
Where is the Sun?
CCD Camera has Discrete Elts

- Lens collects light rays
- CCD elts replace chemicals of film
- Number of elts less than with film (so far)
Camera + Programs = Display

- Camera inputs to frame buffer
- Program can interpret data
- Program can add graphics
- Program can add imagery
Human eye as a spherical camera

- 100M sensing elts in retina
- Rods sense intensity
- Cones sense color
- Fovea has tightly packed elts, more cones
- Periphery has more rods
- Focal length is about 20mm
- Pupil/iris controls light entry

• Eye scans, or saccades to image details on fovea
• 100M sensing cells funnel to 1M optic nerve connections to the brain
Qualitative Representation of Visual Detail of the Eyes


MSU/CSE  Fall 2017
Some image format issues

Spatial resolution; intensity resolution; image file format

MSU/CSE  Fall 2017
Resolution is “pixels per unit of length”

- Resolution decreases by one half in cases at left
- Human faces can be recognized at 64 x 64 pixels per face

MSU/CSE  Fall 2017
Features detected depend on the resolution

- Can tell hearts from diamonds
- Can tell face value
- Generally need 2 pixels across line or small region (such as eye)
Image Super-Resolution

- Scale factor 4 super-resolution results
Super-Resolution Results

Image Processing Operations

Thresholding;
Edge detection;
Motion field computation

MSU/CSE Fall 2017
Find Regions via Thresholding

Region has brighter or darker or redder color, etc.
If pixel > threshold
then pixel = 1 else pixel = 0
Example Red Blood Cell Image

- Many blood cells are separate objects
- Many touch – bad!
- Salt and pepper noise from thresholding
- How useable is this data?
sign = imread('Images/stopSign.jpg','jpg');
red = (sign(:, :, 1)>120) & (sign(:, :, 2)<100) & (sign(:, :, 3)<80);
out = red*200;
imwrite(out, 'Images/stopRed120.jpg', 'jpg')
sign = imread('Images/stopSign.jpg','jpg');
red = (sign(:, :, 1)>120) & (sign(:, :, 2)<100) & (sign(:, :, 3)<80);
out = red*200;
imwrite(out, 'Images/stopRed120.jpg', 'jpg')
Thresholding is Usually Not Trivial

A gray-tone image and the pixels below and above the threshold of 93 (shown in white) found by the Otsu automatic thresholding operator.

MSU/CSE  Fall 2017
Cluster Pixels by Color Similarity and Adjacency

Original RGB Image

Color Clusters by K-Means
Detect Motion via Subtraction

- Constant background
- Moving object
- Produces pixel differences at boundary
- Reveals moving object and its shape

Differences computed over time rather than over space
Two Frames of Aerial Imagery

Video frame $N$ and $N+1$ shows slight movement: most pixels are same, just in different locations.

MSU/CSE  Fall 2017
Best Matching Blocks Between Video Frames N+1 to N (motion vectors)

The bulk of the vectors show the true motion of the airplane taking the pictures. The long vectors are incorrect motion vectors, *but they do work well for compression of image I2!*

Best matches from 2\textsuperscript{nd} to first image shown as vectors overlaid on the 2\textsuperscript{nd} image. (Work by Dina Eldin.)

MSU/CSE  Fall 2017
Gradient from 3x3 Neighborhood

Estimate both magnitude and direction of the edge.

\[
\begin{align*}
\frac{f_y}{f_x} &= \left( \frac{(38-12)/2 + (66-15)/2}{(65-38)/2 + (64-14)/2} ight) / 3 \\
&= \left( \frac{13 + 25 + 11}{13 + 25 + 15} \right) / 3 = 16 \\
\end{align*}
\]

\[
\begin{align*}
\frac{f_x}{f_y} &= \left( \frac{(65-42)/2}{(42-12)/2} \right) / 3 \\
&= \left( \frac{13 + 25 + 15}{13 + 25 + 19} \right) / 3 = 18 \\
\end{align*}
\]

\[\theta = \tan^{-1}(16/18) = 0.727 \text{ rad} = 42 \text{ degrees}\]

\[|\nabla f| = \left( 16^2 + 18^2 \right)^{1/2} = 24\]
2 Rows of Intensity vs Difference
Boundaries not Always Found Well

Blocks image (left) and extracted set of straight line segments (right). The line segments were extracted by the ORT (Object Recognition Toolkit) package.
Canny Edge Operator

Contours from an aerial image of farm fields defined using the Canny operator with $\sigma = 2$ and $\sigma = 1$ respectively. Note that five major structures are well represented – three fields and two straight horizontal bands in the bottom of the image (a canal and a road alongside it).
Color and Shading

- Used heavily in human vision
- Color is a pixel property, making some recognition problems easy
- Visible spectrum for humans is 400nm (blue) to 700 nm (red)
- Machines can “see” much more; ex. X-rays, infrared, radio waves

MSU/CSE  Fall 2017
Imaging Process (review)
Factors that Affect Perception

- **Light:** the spectrum of energy that illuminates the object surface
- **Reflectance:** ratio of reflected light to incoming light
- **Specularity:** highly specular (shiny) vs. matte surface
- **Distance:** distance to the light source
- **Angle:** angle between surface normal and light source
- **Sensitivity:** how sensitive is the sensor
CV: Perceiving 3D from 2D

Many cues from 2D images enable interpretation of the structure of the 3D world producing them
Many 3D Cues

How can humans and other machines reconstruct the 3D nature of a scene from 2D images?

What other world knowledge needs to be added in the process?
What about Models for Recognition

“recognition” = to know again; How does memory store models of faces, rooms, chairs, etc.?
Some Methods: Recognize

- Via geometric alignment: CAD
- Via trained neural net
- Via parts of objects and how they join
- Via the function/behavior of an object
CV Applications and New Topics

MSU/CSE  Fall 2017
Aerial Images & GIS

- Aerial image of Wenatchie River watershed
- Can correspond to map; can inventory snow coverage
Medical Imaging is Critical

- Visible human project at NLM
- Atlas for comparison
- Testbed for methods
Manufacturing case

100 % inspection needed
Quality demanded by major buyer
Assembly line updated for visual inspection well before today’s powerful computers

MSU/CSE  Fall
2015
Simple Hole Counting Alg.

- Customer needs 100% inspection
- About 100 holes
- Big problem if any hole missing
- Implementation in the 70’s
- Alg also good for counting objects
Imaging added to line

- Camera placed above conveyor line
- Back lighting added
- 1D of image from motion of object past the camera

FIG. 2 - LIGHTING SCHEME NO. 1
Critical “corner patterns”

- “external corner” has 3(1)s and 1(0)
- “internal corner” has 3(0)s and 1(1)
- Holes computed from only these patterns!

(a) $2 \times 2$ external corner patterns

(b) $2 \times 2$ internal corner patterns

(c) Three bright holes in dark background
Hole (Object) Counting Alg.

Input a binary image and output the number of holes it contains.

\( M \) is a binary image of \( R \) rows of \( C \) columns.
1 represents material through which light has not passed;
0 represents absence of material indicated by light passing.
Each region of 0s must be 4-connected and all image border pixels must be 1s.
\( E \) is the count of external corners (3 ones and 1 zero)
\( I \) is the count of internal corners (3 zeros and 1 one)

```plaintext
integer procedure Count_Holes(M)
{
    examine entire image, 2 rows at a time;
    count external corners E;
    count internal corners I;
    return(number_of_holes = (E - I)/4);
}
```
\[
\text{#holes} = (\#e - \#i)/4
\]

|   | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0 | 1 | 2 | 3 | 4 | 5 | e | i |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |   |   |
| 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 |   |   |
| 2 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 1 |   |   |
| 3 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 |   | 1 |
| 4 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |   |   |
| 5 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |   |   |
| 6 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |   |   |

(d) Binary input image 7 rows high and 16 columns wide

|   | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0 | 1 | 2 | 3 | 4 | 5 | e | i |
| 0 | e | e | e | e | e | e | e | e |   |   |   |   |   |   |   |   | 6 | 0 |
| 1 |   |   | e | i |   |   |   |   |   |   |   |   |   |   |   |   | 1 | 1 |
| 2 | e | e | e | e | i | e | e | e | i | e | i |   |   |   |   |   | 6 | 2 |
| 3 | e | i | e | i | e |   | i | i |   |   |   |   |   |   |   |   | 2 | 4 |
| 4 | e | i | e | e | e |   |   | e |   |   |   |   |   |   |   |   | 4 | 2 |
| 5 | e | e | e |   |   |   |   |   |   |   |   |   |   |   |   |   | 2 | 0 |
| 6 |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   | 0 | 0 |

(e) External corner patterns marked with e; internal corners marked i.
Variations on Algorithm

- Easy if entire image is in memory
- Only need to have 2 rows in memory at any time
  * used in the 1970’s
  * can allow special hardware
Relate to driving around city blocks.
Check out C++ program and results on web.

MSU/CSE  Fall 2015
Industrial Vision/Inspection

- Literally thousands of applications
- Usually very specific engineering
- Usually called “image processing” – not “computer vision”
Some Hot New Topics

- Phototourism: from hundreds of overlapping images, maybe some from cell phones, construct a 3D textured model of the landmark[s]

- Photo-GPS: From a few cell phone images “the web” tells you where you are located [perhaps using the data as above]

- Deep learning for computer vision
Photo Tourism

http://phototour.cs.washington.edu/
Snap to Paint

Seated Nude (Femme nue assise)' by Pablo Picasso, 1910
"The Shipwreck of the Minotaur" by J.M.W. Turner, 1805
Summary

- Images have many low level features
- Can detect uniform regions and contrast
- Can organize regions and boundaries
- Human vision uses several simultaneous channels: color, edge, motion
- Use of models/knowledge diverse and difficult
- Last 2 issues difficult in computer vision