Computer Vision

George Konidaris
gdk@cs.brown.edu

Fall 2017
sensors

actuators
Image Capture
What’s an Image?
Computer Vision

Image preprocessing

Recognition

Reconstruction

flower

(R&N)
Image Preprocessing

Collection of methods
Typically:
  • Low-level
  • Repetitive
  • Local
  • Easy to parallelize
  • Serve as input to later processing
Edge Detection

(R&N)
What’s an Edge?

“Edges are straight lines or curves in the image space across which there is a “significant” change in image brightness.”

Figure 24.6 Different kinds of edges: (1) depth discontinuities; (2) surface orientation discontinuities; (3) reflectance discontinuities; (4) illumination discontinuities (shadows).
Finding Edges

That gives us a hint!
Compute the derivative of brightness with respect to position.

Brightness:

• Average RGB pixel values:
  \[ Blm(x, y) = \frac{Im(x, y).r + Im(x,y).g + Im(x, y).b}{3} \]

Derivative:

• Take a vertical slice of the image \( H_i = Blm(i, :) \)
• Compute brightness difference between \( H_i(x) \) and \( H_i(x+1) \)
Finding Edges

intensity

R&N
Canny Edge Detector

Classic and very accurate edge detector.

Five steps:

• Gaussian filter to smooth image (reduce noise)
• Find intensity gradients (horizontal, vertical, diagonal)
• Non-maximum suppression
• Threshold to get edges
• Edge tracking: keep only “connected” edges.

https://www.youtube.com/watch?v=xvKr9uruMik
Optical Flow

Useful for understanding movement
Optical Flow

Formally!

Given two images $I_1$ and $I_2$

• Produce optical flow field $F$
  • $F(x, y) = (dx, dy)$
  • where pixel $I_1[x,y]$ moves to $I_2[x + dx, y + dy]$

This boils down to finding correspondences.

One approach

• Find correspondences that minimize “patch” error
• Regularize for smaller movements

https://www.youtube.com/watch?v=GIUDAZLfYhY
https://www.youtube.com/watch?v=_Rpi7WS-HTw
Image Segmentation
Computer Vision

Image preprocessing

Recognition

Reconstruction

(R&N)
Recognition

flower

(ImageNet)
Recognition

Given:
- Object classes $O_1, \ldots, O_n$
- An image size $I$
- A collection of labeled data points $\{i, O_i\}_n$

Find:
- $f : I \rightarrow O_i$

Minimizing expected error.

Classification
Recognition

Why is this hard?

- Foreshortening
- Aspect
- Occlusion
- Deformation
Recognition

Two main ways of going about this:
• Use a geometric object model
• Use machine learning

First: use an object model to *match* an object in a scene.
Recognition by ML

Just do ML:

- Get lots of labeled data
- Learn a classifier

Primary challenge:

- Objects of the same class look different
- The same object looks different from different orientations
Recognition by ML

Solution:

• Compute features from the image
• Features should be invariant to scale, translation, etc.
  • This is a form of *special knowledge* about images.
• Use these as input to classifier instead of image

SIFT features

• Scale-invariant feature transform
• Most widely used
• Many applications in industry
Recognition by Parts

Combine ML and object-models

- Objects are made up of “parts”
- Parts have specific relationships to each other
- Match parts by ML, objects by templates or ML
- Best performing: deformable parts
Deep Nets for Object Recognition

\[ o_1 \rightarrow h_{n1} \rightarrow h_{n2} \rightarrow h_{n3} \rightarrow o_2 \]

\[ ... \]

\[ h_{11} \rightarrow h_{12} \rightarrow h_{13} \]

\[ x_1 \rightarrow x_2 \]

flower
Convolutional Deep Nets

Key idea:
- The first few layers of processing in a deep net construct features automatically.
- Those features should be location invariant.
- Create a layer of neurons with spatially local input.
- Constrain their weights to be the same.
Convolutional Deep Nets
Convolutional Deep Nets

All the usual tricks apply:

- Training vs. test set
- Pretraining
- Can generate synthetic data!
- Must design network architecture
- But no need to think hard about features
- Very powerful hypothesis class
- Lots of data available!

0.23% error rate
Computer Vision

Image preprocessing

Recognition

Reconstruction

(R&N)
Reconstruction

Recover 3D information and structure from collection of images.
Reconstruction

(Tomasi, R&N)
Reconstruction

https://www.youtube.com/watch?v=fLnd9ucUu9Y
https://www.youtube.com/watch?v=EYC6zymrBEE
https://www.youtube.com/watch?v=3BNOsxMZD14
Tracking
Depth Sensors

https://www.youtube.com/watch?v=tlLschoMhuE
3D Perception

Typically given 3D model of *specific* object:
• Identify it from a partial view.
• Pose estimate.
• Complete.

(Glover)
3D Perception in Clutter
3D Perception

Sensing a *novel* chair

true model  observation  reconstruction
Autonomous Cars