CS/ECE 281B – Advanced Computer Vision

- Focus on the DL/AI development in computer vision over the past 5 years
- Though DL/AI activities are in all sub-areas of CV, our focus will be on 3 "D"
  - Detection
  - Description, and
  - Discrimination (Recognition)

Traditional vs. DL/AI frameworks

- Low-level
- Flat
- Hand-crafted
- Separate processes

- High-level
- Hierarchical
- Automated
- Integrated processes

Traditional Detection Mechanism

- Interesting patterns (the question of *where*)
  - Not flat
  - Not Simple
- Harris Corner Detector
- Fast Corner Detector

Description Mechanisms

- (x,y,z) is a descriptor of a 3D vector
- (r, theta) is a descriptor of a 2D vector
- Fourier transform coefficients form a descriptor, so are wavelet coefficients
- Multiple possibilities
- suitable for different tasks

High Dimensional Contents

- Sparsity
- Correlation and redundancy
- The necessity of a paradigm shift
  - description fitting the data rather than
  - data fitting the description
  - Principal Components
  - SVD
  - Dimension Reduction
  - Etc.

Traditional Description Mechanism

- Local patterns (the question of *what*)
  - Chain codes (e.g., Fast)
  - Localized and normalized gradient histograms HOG (e.g., Sift)
- Observations
  - *Where* is relatively simple and uniform
  - *What* is much harder
  - Answers to where and what can be separate processes

Traditional Discrimination Mechanisms

- Not strictly a CV problem, with solutions shared with PR, ML, and AI (e.g., http://www.cs.ucsb.edu/~cs281b/lectureMI.html)
- A rich set of tools
  - Linear discrimination function, logistic regression, support vector machine, decision trees, simple and multi-layer perceptrons, etc.

What can traditional techs do?

- Precise localization and low-level features
- Precision matching (pairing) of them over multiple frames, for accurate
  - Movement
  - Disparity
  - Homography
  - Etc.

Programming assignments

- 2D lip tracking
- Picture stitching
- Stereo reconstruction
- 3D depth inference

What can't traditional techs do?

- The detection/description mechanisms are separated
  - Heard a noise (easy), what kind of noise (party, fight, TV, music, etc.) is hard
  - Need better integration of the 1st 2D (face, pedestrian, vehicle)
- The description mechanism is deficient
  - Hand crafted – limited by human imagination
  - No hierarchy – describing simple patterns

What can't traditional techs do?

- The discrimination mechanisms are separated
  - Integrating discrimination with the first 2D
  - Simpler network structures
  - Faster implementation

This Course

- Convolution
  - For discrete images
  - Generalization of 2D filter (edge, corner, etc.)
- Recurrent
  - For video and time dependent analysis
  - A generalization of Kalman filter
- Advanced topics
  - Discussions by my Ph.D. students
  - Your research presentations and demos

CNN

- Single layer
- Pre-trained weights

- Multi-layers
- Unknown weights (learned from data)
- Nonlinearity a must
- Batch normalization is useful
- Avoid saturation (Resnet and Densenet)

-1	0	1
-2	0	2
-1	0	1

-1	-2	-1
0	0	0
1	2	1

RNN

- Known internal/external (observation) matrices
- Unknown internal state
- Unknown observation matrices
- Avoid saturation (peephole connection, projection, bi-directional, normalization, etc.)

Logistics

- Many variances
  - Tensorflow (Google)
  - Pytorch (Facebook)
  - Mxnet (Apache, Amazon)
- Tensorflow is running in CSIL
  - CPU, not GPU
  - Cannot do much without GPU
- Python encoding only (Matlab? C++?)