Lecture 3
Math & Probability
Background

ch. 1-2 of Machine Vision by Wesley E. Snyder & Hairong Qi

Spring 2016
18-791 (CMU ECE) : 42-735 (CMU BME) : BioE 2630 (Pitt)
Dr. John Galeotti

General notes about the book

- The book is an overview of many concepts
- Top quality design requires:
  - Reading the cited literature
  - Reading more literature
  - Experimentation & validation
Two themes

- Consistency
  - A conceptual tool implemented in many/most algorithms
  - Often must fuse information from many local measurements and prior knowledge to make global conclusions about the image

- Optimization
  - Mathematical mechanism
  - The “workhorse” of machine vision

Image Processing Topics

- Enhancement
- Coding
  - Compression
- Restoration
  - “Fix” an image
  - Requires model of image degradation
- Reconstruction
**Machine Vision Topics**

- **AKA:**
  - Computer vision
  - Image analysis
  - Image understanding

- **Pattern recognition:**
  1. **Measurement of features**
     - Features characterize the image, or some part of it
  2. **Pattern classification**
     - Requires knowledge about the possible classes

---

**Feature measurement**

- **Original Image**
  - **Ch. 6-7** Noise removal
  - **Ch. 8** Segmentation
  - **Ch. 9** Shape Analysis, Consistency Analysis
  - **Ch. 10-11** Matching
  - **Ch. 12-16** Features

---

*Our Focus*
Probability

- Probability of an event $a$ occurring:
  - $Pr(a)$
- Independence
  - $Pr(a)$ does not depend on the outcome of event $b$, and vice-versa
- Joint probability
  - $Pr(a,b) = \text{Prob. of both } a \text{ and } b \text{ occurring}$
- Conditional probability
  - $Pr(a|b) = \text{Prob. of } a \text{ if we already know the outcome of event } b$
  - Read “probability of $a$ given $b$”

Probability for continuously-valued functions

- Probability distribution function:
  $$P(x) = Pr(z < x)$$
- Probability density function:
  $$p(x) = \frac{d}{dx} P(x)$$
  $$\int_{-\infty}^{\infty} p(x) dx = 1$$
Linear algebra

\[ \mathbf{v} = [x_1 \, x_2 \, x_3]^T \quad \mathbf{a}^T \mathbf{b} = \sum_i a_i b_i \quad |\mathbf{x}| = \sqrt{\mathbf{x}^T \mathbf{x}} \]

- Unit vector: \( |\mathbf{x}| = 1 \)
- Orthogonal vectors: \( \mathbf{x}^T \mathbf{y} = 0 \)
- Orthonormal: orthogonal unit vectors
- Inner product of continuous functions
  \[ \langle f(x), g(x) \rangle = \int_a^b f(x) g(x) \, dx \]
  - Orthogonality & orthonormality apply here too

Linear independence

- No one vector is a linear combination of the others
  - \( x_j \neq \sum a_i x_i \) for any \( a_i \) across all \( i \neq j \)
- Any linearly independent set of \( d \) vectors \( \{\mathbf{x}_i = 1 \ldots d\} \)
  is a basis set that spans the space \( \mathbb{R}^d \)
  - Any other vector in \( \mathbb{R}^d \) may be written as a linear combination of \( \{\mathbf{x}_i\} \)
- Often convenient to use orthonormal basis sets
- Projection: if \( \mathbf{y} = \sum \mathbf{a}_i \mathbf{x}_i \) then \( \mathbf{a}_i = \mathbf{y}^T \mathbf{x}_i \)
Linear transforms

- a matrix, denoted e.g. $A$
- Quadratic form:
  $$\mathbf{x}^T A \mathbf{x}$$
  $$\frac{d}{d\mathbf{x}} (\mathbf{x}^T A \mathbf{x}) = (A + A^T) \mathbf{x}$$
- Positive definite:
  - Applies to $A$ if
  $$\mathbf{x}^T A \mathbf{x} > 0 \quad \forall \mathbf{x} \in \mathbb{R}^d, \mathbf{x} \neq 0$$

More derivatives

- Of a scalar function of $\mathbf{x}$:
  - Called the gradient
  - Really important!
  $$\frac{df}{d\mathbf{x}} = \left[ \frac{\partial f}{\partial x_1} \frac{\partial f}{\partial x_2} \ldots \frac{\partial f}{\partial x_d} \right]^T$$
- Of a vector function of $\mathbf{x}$
  - Called the Jacobian
  - Hessian = matrix of 2nd derivatives of a scalar function
Misc. linear algebra

- Derivative operators
- Eigenvalues & eigenvectors
  - Translates “most important vectors”
  - Of a linear transform (e.g., the matrix \( A \))
  - Characteristic equation: \( Ax = \lambda x \quad \lambda \in \mathbb{R} \)
  - \( A \) maps \( x \) onto itself with only a change in length
  - \( \lambda \) is an eigenvalue
  - \( x \) is its corresponding eigenvector

Function minimization

- Find the vector \( x \) which produces a minimum of some function \( f(x) \)
  - \( x \) is a parameter vector
  - \( f(x) \) is a scalar function of \( x \)
    - The “objective function”
  - The minimum value of \( f \) is denoted:
    \[ \hat{f}(x) = \min_{x} f(x) \]
  - The minimizing value of \( x \) is denoted:
    \[ \hat{x} = \arg\min_{x} f(x) \]
Numerical minimization

- Gradient descent
  - The derivative points away from the minimum
  - Take small steps, each one in the “down-hill” direction
- Local vs. global minima
- Combinatorial optimization:
  - Use simulated annealing
- Image optimization:
  - Use mean field annealing

Markov models

- For temporal processes:
  - The probability of something happening is dependent on a thing that just recently happened.
- For spatial processes
  - The probability of something being in a certain state is dependent on the state of something nearby.
  - Example: The value of a pixel is dependent on the values of its neighboring pixels.
Markov chain

- Simplest Markov model
- Example: symbols transmitted one at a time
  - What is the probability that the next symbol will be \( w \)?
- For a “simple” (i.e. first order) Markov chain:
  - “The probability conditioned on all of history is identical to the probability conditioned on the last symbol received.”

Hidden Markov models (HMMs)
HMM switching

- Governed by a finite state machine (FSM)

The HMM Task

- Given only the output $f(t)$, determine:
  1. The most likely state sequence of the switching FSM
     - Use the Viterbi algorithm (much better than brute force)
     - Computational Complexity of:
       - Viterbi: $(\text{# state values})^2 \times (\text{# state changes})$
       - Brute force: $(\text{# state values})^{\text{# state changes}}$
  2. The parameters of each hidden Markov model
     - Use the iterative process in the book
     - Better, use someone else's debugged code that they've shared