CS370: Computer Vision. Review for final

Erik G. Learned-Miller
Department of Computer Science
University of Massachusetts, Amherst
Amherst, MA 01003

May 2, 2011

Abstract

The final will cover material from the entire term. Focus particularly on the documents posted on the course web page. Of course, there may be material from class presented as well.

- Basic math:
  - Dot products. Consider not only the dot product of two vectors, but the dot product of two image patches. This is just the sum of the product of corresponding pixels.
  - Partial derivative of an image "function" in the horizontal direction and vertical directions. This can be approximated by just taking the difference between neighboring pixels in either the horizontal direction, or the vertical direction (see section on convolution below).
  - Gradient. The gradient is simply a vector formed by stacking up partial derivatives. The gradient of an image at a particular point is just

    $$\begin{vmatrix} \frac{\partial f}{\partial x} \\ \\ \frac{\partial f}{\partial y} \end{vmatrix}$$

    I can think of the direction of the gradient as the angle it makes with the x axis. Then, if I want, I can form a histogram of these directions at different places in the image, leading to a histogram of gradients. Such histograms of gradient directions are sometimes referred to as HoG features.
  - L1 distance, L2 distance (Euclidean distance), L_p distance.
  - Marginalization, conditional probabilities, joint probabilities. Bayes rule.
  - Convolution (see entire section on convolution below).
  - Associativity. Associativity is the property of an operator that one can "regroup" operations with parentheses leaving the results unchanged. It does NOT imply a reordering of the operations, which is known as commutativity. Example of associativity:

    $$a + (b + c) = (a + b) + c.$$

    Note that subtraction is NOT associative:

    $$a - (b - c) \neq (a - b) - c,$$

    in general. Note that **convolution** is associative. This means that it is convenient to apply multiple convolutions to an image because we can write

    $$A * (B * Image) = (A * B) * Image$$

    meaning we can perform a convolution of two operators A and B ahead of time and apply a single convolution to the entire image rather than applying the operators one after another to the image, which is about twice as expensive.

- Review Bayes rule thoroughly. It is the single most important thing from this course. You should be able to
  - Write down the math.
  - Understand the meaning of likelihoods, prior probabilities ("priors"), and posterior probabilities ("posteriors").
  - Compute the denominator in Bayes rule by writing down a marginalization and the expansion of the joint probability in terms of the conditional probabilities that you have access to.
  - Most importantly, you should understand how to apply Bayes rule. Two big examples include classification of handwritten digits and classification of foreground and background pixels in video.
- You should know everything in the "Light sources and camera models" handout on the course web page.
  - Understand the difference between a point light source and an extended light source. Note that point light sources don't really exist (since a light source has some finite positive size.) They are just a convenient abstraction.
  - Understand the inverse square law, *and the basic geometry behind it.*
  - Understand the concept of "solid angle" and the units of steradians. Understand how they relate to the radians in a circle. How many steradians are in a sphere? Half a sphere? A square meter of solar panel (given the distance to the sun).
  - What is the basic purpose of a lens? What advantages does a camera with a lens have over a pinhole camera? What disadvantages?
- A key part of computer vision is building probability models from data. For example, when doing background subtraction, we may want to record the brightness values we see at a particular pixel for some time and define a "background model", which gives us the probability $p(x|b)$ of seeing a particular brightness value x at a given pixel, given that it came from the background. You should understand the following issues related to this type of modeling.
  - A simple way to build such a model is to simply collect a histogram of how many times each brightness value is seen, and then divide by the number of samples to produce a probabilty distribution.
  - What is the problem with this naive approach? Answer: It might take a very large number of samples to get an accurate view of the probability distribution.
  - How can we mitigate this problem? Answer: We can "smooth" the distribution by spreading counts across multiple bins.

  - How can we smooth counts across bins in a simple and efficient manner? Answer: Convolve our naively obtained distribution with a "kernel" that spreads the data out. (See convolution below.)
  - Another way to model a background is simply to compute the mean value of the observed background values. Values near the mean can then be considered to be part of the background. What is a fundamental probelm with such a method? Answer: If there are two brightness values present in a particular location (caused by a waving tree for example), then the mean does not characterize either of these values well.

- Understand the following about the process of convolution.
  - Convolution takes an image and a convolution kernel, which can be thought of as a mini-image, and produces another image by replacing each pixel in the original image with the dot product of 1) a patch surrounding that image, and 2) the convolution kernel.
  - Be able to compute, by hand, the convolution of a simple discrete function with a simple convolution kernel. To do this, you should be able to compute the dot product between two mini-images, and to understand how to "slide" the convolution kernel across each spot in an image to compute this dot product at each location.
  - Understand how convolution relates to
    - Increasing the aperture (hole) size in a pinhole camera. Answer: Instead of obtaining a sharp image as with a traditional pinhole camera, it results in a blurry image which can be seen by convoling the ideal pinhole camera image with a convolution kernel defined by the shape and size of the pinhole aperture.
    - Taking a picture of a moving object. Answer: Taking a picture of a moving object with a long exposure (period during which light is recorded) is like taking a sequence of pictures of a moving object and averaging them all together. This results in the object being blurred in the direction of motion. This is exactly what happens if you convolve a still image with a mini-image that represents the path of the object with respect to the camera.
    - The blurryness of an image of the stars. Answer: Blurry astronomical images result from the light from very small sources (stars) being perturbed by the earth's atmosphere, like in a Pachinko machine. This can be modeled by taking the "true" image of the stars and convolving it with a small normal (or Gaussian) distribution.
    - Edges and image derivatives. Answer: Edges in images are related to the difference in adjacent brightness values. If we want to find vertical edges in an image, this is essentially an attempt to find the difference between each pixel and its neighbors on

the sides. This can be modeled by taking the difference between each pixel and the pixel on its right. Alternatively, it can be modeled by taking the value on the left of a pixel and subtracting the value on its right. Either of these can be represented as a mini-image (the first by $[1 \quad -1]$ and the second by $[1 \;\; 0 \quad -1]$) and the application of a convolution.

- Face recognition.
  - What are the major subtasks of face processing? Answer: Detection, alignment, and recognition.
  - What does the “face in the beans” image suggest about human face detection?
  - Discuss why alignment is important for face recognition.
  - Discuss the problem of learning to recognize whether two martians are the same from pairs of pictures of martians that are the same or different. How does this relate to the problem of recognizing whether two humans are the same or different? Discuss the properities of good “features” for recognition. Answer: features must be *repeatable* and *discriminative.*