Homework 1 (due Monday Jan 24 by end of the day)

1. Consider the (unnormalized) 2D bivariate distribution f(x,y)

	x=1	x=2	x=3	x=4
y=0:	10	10	10	10
y=1:	10	20	20	0
y=2:	0	10	0	0

For each of the following, give your answer as a NORMALIZED distribution (sums to 1)

a) What is the marginal distribution g(x) ?
b) What is the marginal distribution h(y) ?
c) What is the conditional distribution p(x | y=0)?
d) What is the conditional distribution p(y | x=2)?
e) Do you think that random variables x and y are independent?
f) What is the expected value of x?
g) What is the expected value of sqrt(y)?

2a. Let f(x,y) be a general bivariate distribution, either discrete or continuous, and let g(x) be its marginal distribution with respect to x. Prove that the expected value of x with respect to f(x,y) is equal to the expected value of x with respect to g(x). That is, we want to show that the joint mean of x is equal to the marginal mean of x.

2b. Come up with a simple counterexample of a discrete distribution f(x,y) and its marginal distribution g(x) to show that the x coordinate of the MODE of f(x,y) is not necessarily the same as the mode of the marginal g(x). Recall that the mode of a discrete distribution is the location where the highest probability value p_max occurs, assuming that value is unique (there are alternate definitions if the distribution has more than one mode... choose your example so that f(x,y) and g(x) each have only one mode).

3a. Explain how you might use or modify NCC score to define a 2D probability distribution describing likely locations of an intensity template in a new image? hint: Are there any problems with using an array of NCC scores as an unnormalized distribution?

3b. Repeat the above question, but using SSD score rather than NCC score

4. Assume we have a 3D volumetric dataset V containing greyscale intensity values (it might be a medical CT scan, for example), and let's generalize the notion of integral image to "integral volume", where any voxel at location X,Y,Z contains the sum over all voxel values of V within the region defined by x<=X, y<=Y,z<=Z. Generalizing from the formula for computing the sum inside a 2D rectangular region using a 2D integral image, tell me how you would compute the sum inside a 3D rectangular parallelepiped (cuboid) region xlo<=x<=xhigh, ylo<=y<=yhigh, zlo<=z<=zhigh using the 3D integral volume data structure.

5. Programming: Choose some images that "look very different" to you with regard to their overall color "palette", and compute a set of three empirical bivariate distributions for each, one representing the joint distribution of red and green, the second representing the distribution of green and blue, and the third representing the distribution of red and blue. Display these either as intensity images, or using matlab's imcontour function to plot contours of the distribution.

Note1: these will be discrete distributions, represented by either normalized or unnormalized histograms. If you use raw R,G,B color values, each histogram would be a 256x256 array of numbers. However, to save computation time (and to get "smoother" estimates... we will talk about this later in the course), you may want to quantize each color dimension to less than 256 values. For example, the upper left element of a 16x16 Red-Green histogram would store the count of pixels where 0<=R<=15 AND 0<=G<=15.

Note2: a fun tool for finding images with extreme color combinations can be found at http://labs.ideeinc.com/multicolr/