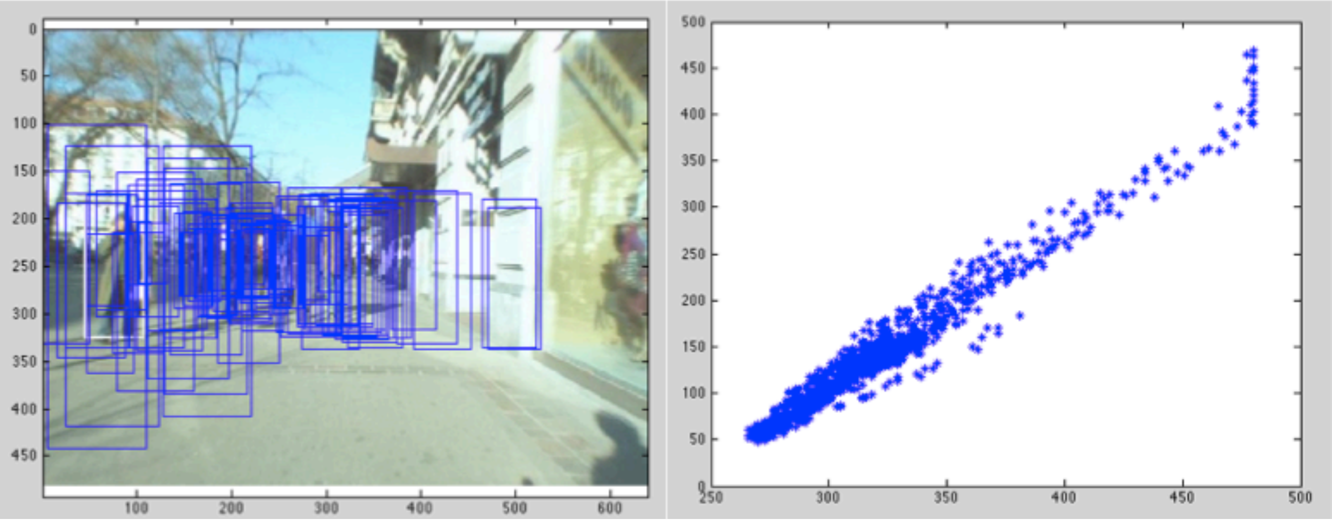
EE554/CSE586 Spring 2016 Assignment 1 Due Date: Thursday, Jan 21, 9:00AM

We have discussed in class why there seems to be a linear, statistical relationship between foot location of a person in an image and the height of their bounding box, at least for typical elevated camera views. In this assignment we will use a discriminative regression method to learn a statistical model mapping foot location to the mean and variance of a normal distribution describing bounding box height expected at that location. We can think of this as a data-driven form of camera calibration, and it is useful because it allows us to do things like filter out incorrect responses from a pedestrian detector when they are not consistent with the geometry of the scene (bounding box too large or too small to feasibly be a person at that location).

We will use data from the multiple object tracking benchmark found at https://motchallenge.net. I have already downloaded and placed the annotation data they provide on our Angel site. For several small video clips, the annotators have meticulously outlined bounding boxes around all the people. Following the nomenclature in Chapter 6 of the textbook, we will use this ground truth as our training data pairs (xi, wi), with xi being the observed "data" value of foot Y (row) location in the image, and wi being the predicted "world" value of bounding box height in the image. I have also provided some sample code (called samplecode.m) that allows you to choose one of the video datasets, load the ground truth data and a sample image from that camera view, display a random sampling of some of the bounding boxes overlaid on the image, and plot all the (xi, wi) foot vs height pairs. One of the datasets is illustrated below. In the plot, x axis is xi values, y axis is wi values.



Following the regression model in Chapter 6 and discussed in class, assume that $P(wi \mid xi, \theta) = \text{Norm}(\phi_0 + \phi_1\, xi, \sigma^2)$ for i = 1,2,…,N where N is the number of training data pairs.

1) Assuming data pairs are independent and identically distributed, write down the joint likelihood function of the parameters $\theta = \{\phi_0, \phi_1, \sigma^2\}$ for the observed data pairs { (xi,wi) ; i=1,…N }.

2) Following the principle of maximum likelihood estimation, derive the values of $\{\phi_0, \phi_1, \sigma^2\}$ that maximize the above joint likelihood. It may be helpful to refer to section 4.4.1 in the text, which works through a simpler example of estimating mean μ and variance σ^2 of a normal distribution from a set of points. In our case, mean is a function of two parameters $\phi_0 + \phi_1$ rather than being a single parameter μ. Nonetheless, the basic solution strategy will be the same: we can show that ϕ_0 and ϕ_1

can be solved for first, without σ^2 , and then σ^2 can be found using the previously estimated values of ϕ_0 and ϕ_1 . Since ϕ_0 and ϕ_1 are coupled, you won't be able to solve one without the other (i.e. you have to solve them jointly). In this case, derive the matrix equation that represents the linear system of equations that needs to be solved to determine the two values. That is, show what 2x2 matrix A and 2x1 matrix b of A x = b are, where x is the 2x1 vector containing parameters ϕ_0 and ϕ_1 .

3) Implement the above mathematical results as code to compute $\{ \phi_0 , \phi_1 , \sigma^2\}$ from (xi, wi) training pairs extracted from the datasets provided. Run it on at least the following two datasets, "ETH-Sunnyday" and "PETS09-S2L1", and report the results. There are two things to consider here. First, is that you might want to remove points from the dataset corresponding to bounding boxes that touch or even exceed the very bottom row of the image, as these lead to corrupted measurements. We can see this in the 2D plot shown earlier – the points at the far right do not follow the linear relationship that we are trying to infer, and if not removed they may corrupt the estimated results. The second thing to think about is how are you going to show your results in a way that convinces me you have the correct answer? I do want you to report the numerical values you estimated for ϕ_0 , ϕ_1 and σ^2, which I will compare with my own, but also show some visualizations that convince me that you got the right answers. One idea is to overlay on the plotted (xi, wi) points the line representing the computed mean relationship wi = $\phi_0 + \phi_1$ xi, and perhaps show as two other lines the confidence intervals for plus or minus two or three standard deviations. Another idea, more practical for a vision person, is to overlay on the sample image frame a set of bounding boxes that you compute for a coarse grid of candidate foot locations, e.g. every 200th row and col location (make sure you only use rows below the horizon line of the image, and also stagger the grid so boxes don't all line up vertically) using the mean value for height computed for a foot at that location. To do this, you will also need to figure out what box width to use for each location – the easiest thing to do is to use a constant proportion of the computed height, but perhaps you could also learn width as a function of foot location from the training data in a similar way to how you learned the height.

What to hand in: Upload in the Angel dropbox a zip file containing your code, and a pdf that shows your math derivations for parts 1&2 and your results/visualizations for part 3.