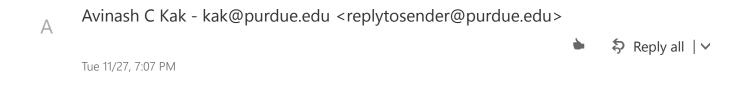


Fall-2018-ECE-66100-001: about the Viola and Jones face detector



Inbox

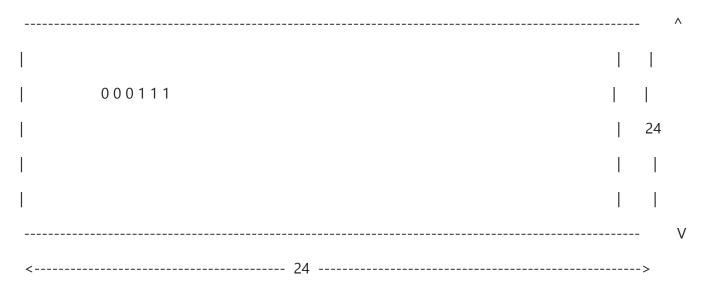
Hello Class:

Based on the questions I was asked after class today, I think many of you may still be confused about how each of the 160,000 features is meant to be used as a weak classifier for a 24x24 window.

Consider the following operator that we wish to use a feature:

0 0 0 1 1 1

at the following EXACT placement in a 24x24 window:



At the SPECIFC LOCATION where I have shown the feature operator in the 24x24 window, it will produce a value between -510 and +510 for an 8-bit image. To find the value produced by the feature operator, you add the pixel values where you see 1's in the operator and subtract from that sum what you get by adding the pixel values where you see 0's in the operator.

The important thing to realize is that the value of the operator AT THE SPECIFIC LOCATION I have shown above in the 24x24 window is meant to be used as a weak classifier for face detection. We find the value of this feature for all the training data images.

Let us say we have a total of N images in the training dataset. WE SORT THE IMAGES according to the value they produce for the feature I have shown above. Now we walk through all possible values for this feature, from the lowest to the highest, and, treating each value as a possible decision threshold, we find



polarity when the images above the threshold are declared as face images and those below the threshold as non-face images. And we have another polarity when the images below the threshold are declared as face images and those above the threshold as non-face images. For the feature shown above, we choose that threshold \theta and that polarity p which yields the lowest misclassification rate. We can now characterize the detection power of this feature by the triple (f, \theta, p) where f is a symbolic name for the feature, \theta the decision threshold that yielded the lowest misclassification error, and p the polarity.

We repeat the steps described above for each of the 160,000 features. At each iteration of the AdaBoost loop, we choose the triple (f,\theta,p) which gives us the best misclassification rate.

Avi