CSCI 5561 Assignment 1

Luis Guzman (guzma102@umn.edu)
September 25th, 2020

In this assignment, I built a Histogram of Oriented Gradients (HOG) descriptor for the purpose of facial recognition within images. The HOG is computed by first filtering the image with the x and y Sobel filters, and computing the image gradient's magnitude and direction from the filtered images:

$$|\nabla I(i,j)| = \sqrt{I_x(i,j)^2 + I_y(i,j)^2} \qquad \angle I(i,j) = \text{atan2}\left(I_y(i,j)^2, I_x(i,j)^2\right)$$

where I_x and I_y refer to the x and y Sobel filtered images, respectively. The negative gradient angles are rotated so that they are within the range $[0, \pi]$. The histogram is then computed by splitting the image into 8-pixel-wide cells, and adding each pixel's gradient value to one of six bins, corresponding to a range of gradient angles. The result of this build_histogram function is an (M x N x 6) matrix, where M and N are the number of cells in the image. Next, I performed block normalization to make the algorithm more robust to changes in illumination and contrast. This aspect consisted of taking each 2x2 block of cells, concatenating the histogram into one long vector, and normalizing that vector's length.

The HOG descriptor can then be used to detect faces by computing the HOG for a template face image and for each pixel in the target image. By comparing the correlation of each HOG vector, we know how much each part of the image resembles a face. Lastly, I used non-maximum suppression (NMS) to delete the extraneous bounding boxes, so that we're left with one box around each face.
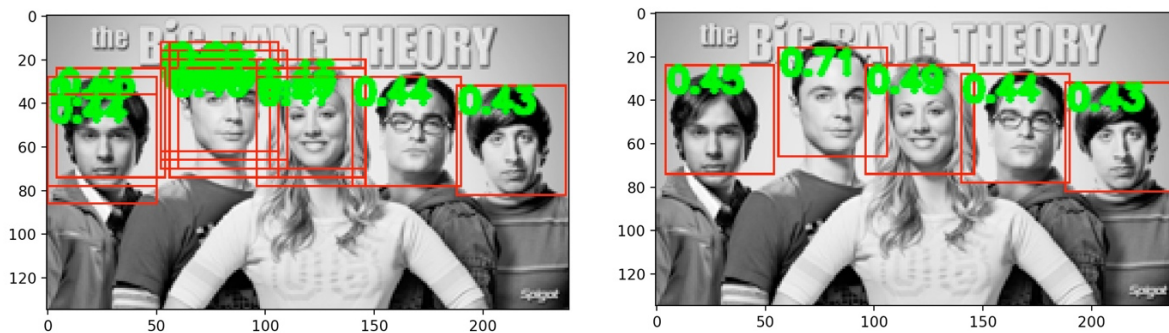

Figure 1: The facial detection before and after NMS (stride value of 4)

My resulting program is able to reliably detect all faces within the target image. Additionally, the HOG descriptor (below) can be visibly checked for correctness since the histograms follow the edges within the image. I ran into issues with the program taking significant time to run, so I defined a stride value for face detection. I found that a stride value of 4 resulted in the best trade off of runtime and number of bounding boxes found. Furthermore, changing the stride value may affect what bounding box threshold is required, since there is more possibility for errors. Further work could include optimizing the code so that a stride of 1 could be used.
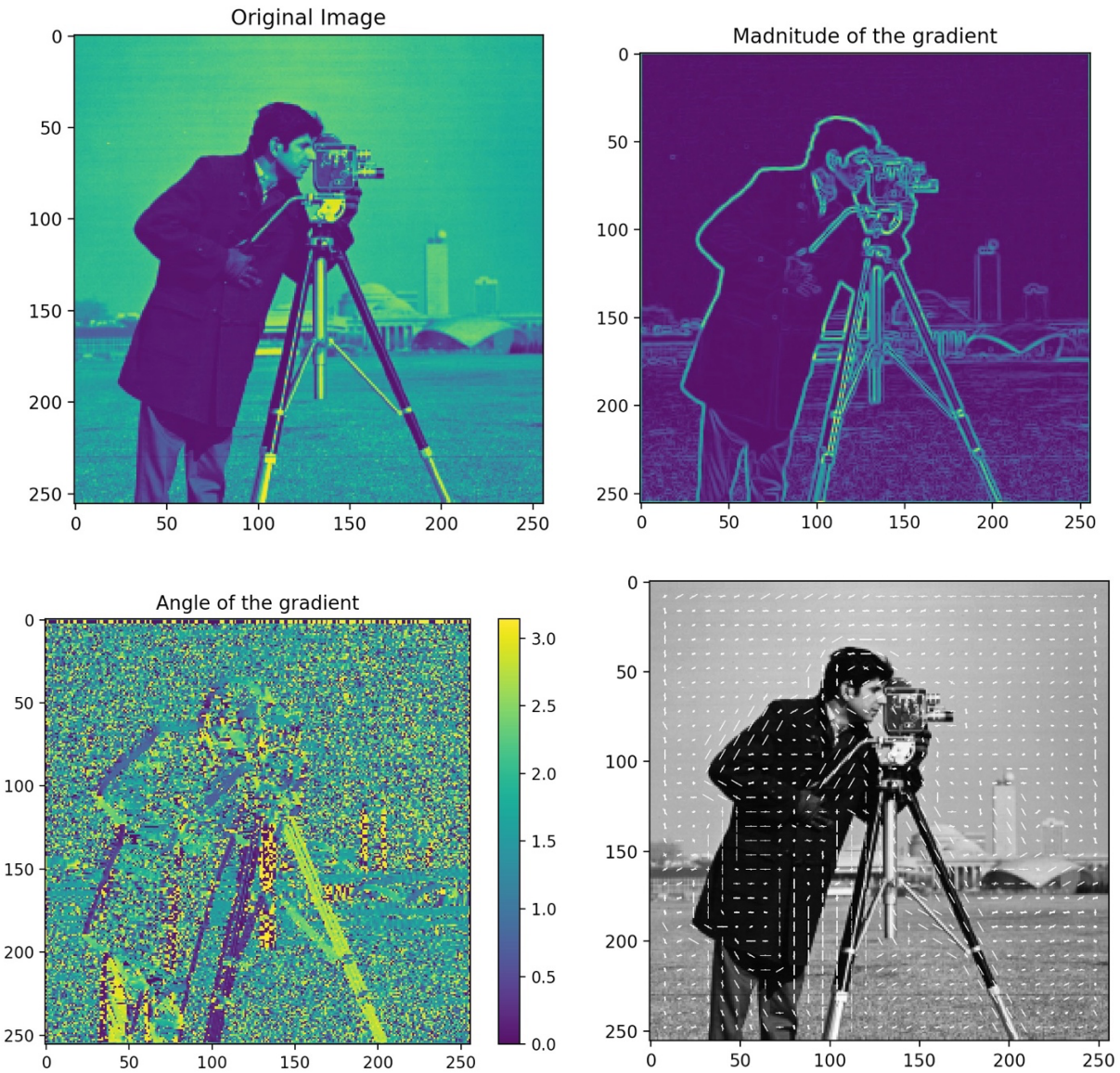
Figure 2: The original, gradient magnitude, gradient angle, and HOG descriptor for the cameraman image.