

CSCI 5561: Image Processing, Fall 2021

Homework 1 Submission

Ashvin Pidaparti, 5421833

Summary

Object detection with Computer Vision has been a long standing problem. Prior to the popularity of Deep Learning, one method has prevailed- a Histogram of Oriented Gradients. In short, the input image's gradients are calculated, and the distribution of their orientation is constructed in a histogram in blocks, in which the magnitude of each is summed with the magnitude of other pixels within the corresponding bucket of orientations. The normalized dot product is computed, yielding a normalized cross-correlation score between a bounding box with the dimensions of a template image and the input image. Should this score be below a threshold value, the box is rejected. This process continues across the entire input image, yielding several bounding boxes. Between bounding boxes above that threshold, a non-maximum suppression is performed. This takes every pair of bounding boxes, determines the area of overlap between the two and the total area covered by both boxes (counting the overlap area only once), and returns the overlap area divided by the total area. If this value is above a second threshold (.5, provided by Dr. Park), we can be reasonably certain that the algorithm has detected 2 matches for the template image that are actually the same object in the input image, thus we discard the bounding box with the lower normalized cross-correlation score. The results are shown below, including a visualization of an image's gradient, a template image, and an input image the algorithm matched to the template image.

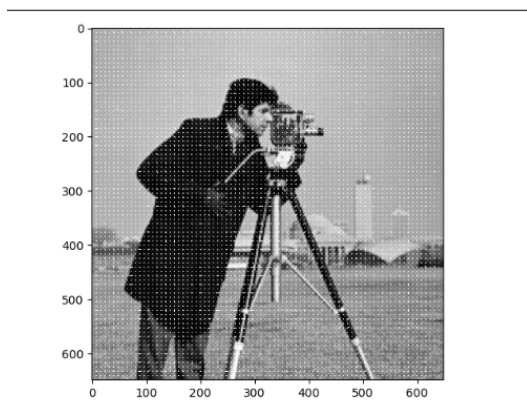


Fig. 1. Image gradients visualized



Fig. 2. Template image used to cross-correlate with input

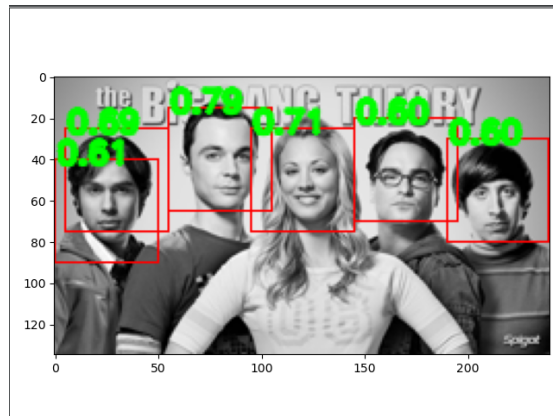


Fig. 3. Input image with faces detected