The Dartboard Challenge: Report
Image Processing and Computer Vision
COMS30121

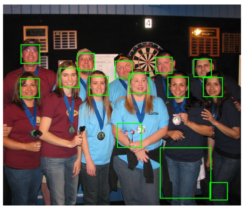
Faizaan Sakib
User ID: ss16161

1 The Viola-Jones Object Detector

1.1 Face Detection

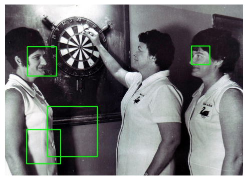
The provided Viola Jones detector for faces was tested with 5 different images that featured faces within the frame. Below are the resulting images showing the faces that were detected with the use of green bounding boxes.


(a) dart4.jpg


(b) dart5.jpg


(c) dart13.jpg


(d) dart14.jpg


(e) dart15.jpg

1.2 Evaluation of the Face Detector

For the image `dart5`, the True Positive Rate (TPR) was 100%, as all 11 valid faces within the frame were detected. On the other hand, `dart15` is an anomaly in the sense that, given my own discretion, has no valid faces within it. As a result, it is not possible to determine the performance on this image.

As mentioned above, the decision to choose what counts as a true positive and what does not is up for interpretation. There is the case of deciding the validity of faces present within the images, which leads to a bound of how well the detector is expected to perform. Also, one has to decide the constraints of what can be permitted as a true positive of a valid face.

Being too aggressive or too lenient with these decisions can end up giving a false representation of how well the detector performs. Therefore, it is important to understand the detector's capabilities and being consistent with the expectations of what it can achieve. `dart15` is a good example of how important these considerations are.

It is possible for a detector have over-fitting tendencies. This can easily lead to correct detections, but with a surplus of incorrect ones. However, the TPR measure does not take into account false positives. This means a detector that finds 10 faces on an image with only 1 face would have the same TPR as one that perfectly finds the 1 valid face and no more. Because of this, the TPR is not the best measure for evaluating a detector's performance.

Instead, the F1 score can be seen as a more viable measure. It is calculated using the precision, which is the proportion of true positives to all positives, and the recall (TPR). These are useful measures, but put together as a score allows comparison between detector performance.

In order to ensure accurate and meaningful calculations of the F1 score, I decided to define my constraints within the generated boxes holding ground truth. Firstly, a face will only be chosen to be valid if it is a frontal and upright human face, with both eyes, nose and mouth being visible, as per the trained classifier. This is why `dart15` does not include any valid faces. Secondly, the bounding box must cover at least the eyes and the mouth, and must intersect with at least 60% of the area of its corresponding ground truth bounding box.

2 Building & Testing Our Own Detector

2.1 TPR vs FPR During Training

During the boosting procedure, the classifier was built in 3 stages. Having collected the TPR and FPR for each of the stages, we can get an insight into how the detector is trained. Starting from the first stage, both the TPR and FPR are at 1. This shows, that the desired initialisation of the classifier, is to find at least every face. It does this by starting off with a very low sensitivity threshold, making the classifier detect any and all objects within the frame. The TPR remains at 1 at all stages, but by the second stage, the FPR is reduced by 96%, instantly reducing the number of false positives by a very large amount. This is because at this stage, a weak classifier is used to immediately discard many false positive sub-windows. Moving on from this, to reduce the FPR even further by stage 2 as it does, more complex computations are required as the FPR is already very low. The attentional cascade builds on the classifier on every stage, introducing complex classifiers or a combination of existing weak classifiers to focus on sub-windows where there is a higher probability of containing a dartboard while further discarding false negatives.

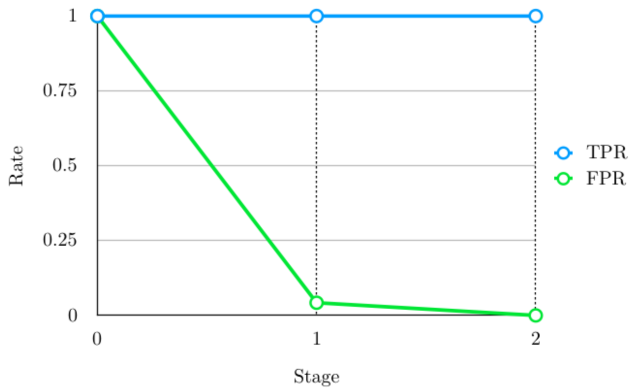
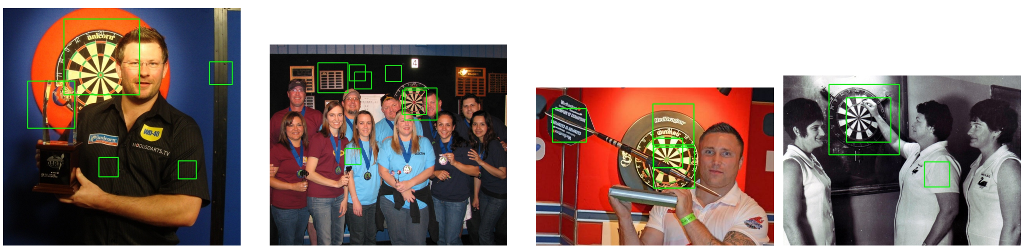


Figure 2: A scatter plot showing the TPR vs FPR of the dartboard detector in each stage of the classifier training.

2.2 Evaluation of the Dartboard Detector

After testing the dartboard detector on the given images, it scores an average F1 score of 0.565. It seems to achieve varying results, understandably coming from the variety of images used. The detector performs well for images where the dartboard appears large on the frame and is lit well under conditions. On images where the dartboards shown are partially covered or appearing differently due to the viewing angle, the detector still does very well, although getting multiple detections for each one. Its performance on `dart11` was notable, detecting the very small segment of dartboard showing to the right. Its performance is similar to the classifier, in that it's very good at getting a TPR of 100% (apart from `dart10`). However it seems to have a much higher FPR than the classifier.



(a) dart4.jpg (b) dart5.jpg (c) dart13.jpg (d) dart15.jpg

Image	dart0	dart1	dart2	dart3	dart4	dart5	dart6	dart7	dart8
F1 Score	0.5	1.333	0.2	0.5	0.333	0.5	1	0.235	0.375

Image	dart9	dart10	dart11	dart12	dart13	dart14	dart15	Average
F1 Score	0.333	0.417	0.667	1	0.4	0.25	1	0.565

3 Integration with Shape Detectors

3.1 Performance of Implementation on Example Images

The images below show how the detector is capable of detecting dartboards that are clearly in sight. However, it fails in extreme cases, where the dartboard is very far away and distorted due to viewing angles.

Figure 4: Left to right: Thresholded gradient magnitude, 2D representation of the Hough Space and the final detection image for `dart3` and `dart8` respectively.

3.2 Evaluation of the Detector using Hough Transform

The average F1 score for the new detector throughout saw an increase from 0.565 to 0.75. For a lot of the images, the F1 score has converged to either a 0 or a 1. The precision was often either 0 or 1, getting perfect results, or not detecting anything. Meanwhile, recall saw a slight decrease too, simply because fewer true positives were being detected.

Image	dart0	dart1	dart2	dart3	dart4	dart5	dart6	dart7	dart8
F1 Score	1	1.333	1	1	1	1	1.333	0	1

Image	dart9	dart10	dart11	dart12	dart13	dart14	dart15	Average
F1 Score	1.333	0	0	0	0	1	1	0.75

Key points:

- Dartboards in clear view are detected very well. Detected lines came with accurate length and angle information.
- Some partially obstructed dartboards can be detected.
- False positives were greatly reduced by looking for lines crossing in the centre of the box.
- Detector's criteria of only detecting with high confidence meant that, although most false positives would be removed, it also meant true positives would also at times be discarded.

3.3 Combining Viola Jones and the Hough Transform

- Detections from the Viola Jones detector were used as the starting point, as it had a high TPR.
- The boxes were cropped from the input, to get better results for the magnitude & Hough Space.
- Hough Line transform was used to detect lines in these boxes. To look for possible intersections, the box was scaled down by 75%, before checking if the midpoint of a line existing within the space.
- To reduce FPR and obtain true positives, a box with greater than 4 midpoints in the search space would be considered a dartboard.

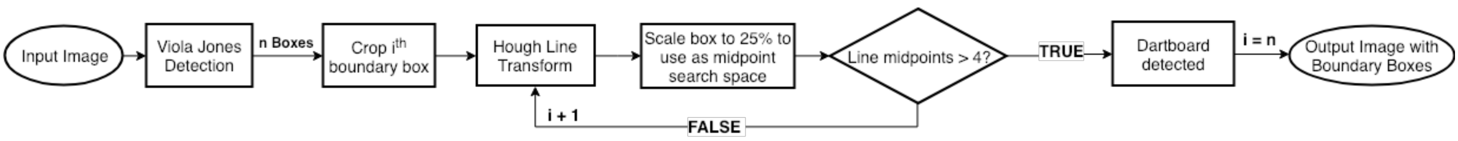


Figure 5: Flow diagram showing how Viola Jones and the Hough Transform was combined. Note: After the completion of sub-task 4, this flow alters to include the consideration of detected circles and the merging of overlapping true positives.

4 Improving the Detector

4.1 Hough Circles

Having implemented the Hough Line Transform, the next logical step would be to use the information that dartboards are categorically circular. A number of dartboards were missed by the previous detector as sometimes the line's midpoints would very narrowly miss the search space; a bottleneck introduced by the Viola Jones detection to start off with. Using OpenCV's `HoughCircles()` function, along with some parameter adjustments, it was accurately detecting circles in images. It also seemed to detect circular properties in the elliptical dartboards that are at an angle. This allowed me to greatly improve the TPR, by saying a box has a dartboard if there are more than 5 line midpoints ***or*** if there is at least 1 circle centre ***and*** 1 line in the reduced search space. This gave more power to either feature dynamically. As seen on Figure 6, the detected circle boosts the confidence on a dartboard on which only 2 lines were detected.

Figure 6: Circle detected on `dart6`

4.2 Merging True Positives

By this point, the detector had become very good at finding true positives. However, the recall could still be improved on, as a dartboard may be detected twice, since the Viola Jones detector was prone to doing this. Thus, I made the detector merge any true positive boxes that overlapped by more than 80%, by taking an average of their position and scale. On Figure 7, there would be 3 boxes on the same dartboard in the image `dart11`. With the help of merging, it reduces any unnecessary true positives, but also can lead to a more accurate placing of the boundary box.

Figure 7: Merging of true positives on `dart11`

4.3 Foreground Extraction using the GrabCut Algorithm

While going through the documentation of OpenCV, I found the function `grabCut()`, which is an implementation of the segmentation algorithm of the same name. It is used to extract the foreground from the background of an image given a bounding box. It iteratively clusters pixels into either category by specifying a weight for each pixel edge, where an edge has a larger weight if the two pixels connected by the edge are of a similar colour, and vice-versa. Although not explicitly improving the performance of the detector, it does help in visualising the goal of the detector.

4.4 Evaluation

After spending time on improving the detector and testing it, here are my findings on the final implementation:

- Overall average F1 score has further increased, due to increase in precision and decrease in recall.
- Addition of circles leads to a higher precision as dartboards that weren't detected before now are.
- Recall is reduced by merging true positives. This has led to better placement of boundary boxes.
- Generally, further away/skewed dartboards are being recognised.
- It is still possible for noise in an image to produce a false positive.
- This detector is still heavily dependent on the results of the Viola Jones detector. As such, some boundary boxes, although correct, may not be very precise.

Image	dart0	dart1	dart2	dart3	dart4	dart5	dart6	dart7	dart8
F1 Score	1	1	1	1	0.667	1	1	0.667	0.667

Image	dart9	dart10	dart11	dart12	dart13	dart14	dart15	Average
F1 Score	1	0.8	1	1	0	0.8	1	0.907