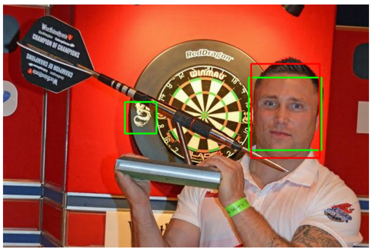
CW1

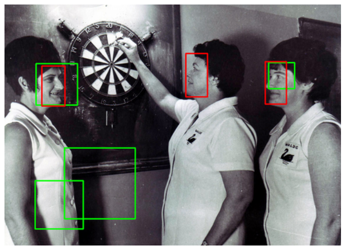
Min Nguyen & Charana Nandasena {mn15104, an15739}

Section 1 – Viola Jones Face Detector Analysis



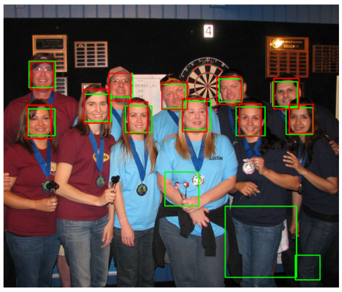
TPR: 1.00
F1: 0.67

TPR: 1.00
F1: 0.50



TPR: 0.67
F1: 0.57

TPR: 1.00
F1: 1.00



TPR : 1.00
F1: 0.88

Viola detections can be seen in green and our annotated bounding boxes in red. Our true positive rates were determined by matching each annotation with its viola detection with highest overlap percentage (of 80% of higher).

1.1 – Challenges of TPR Assessment

The difficulties in assessing TPR accuracy in this context involves the question, what should we perceive a frontal face to be? Under the assumption that the classifier was trained purely on non-occluded, face-on perspectives of faces, the TPR rate produced was as accurate as possible. From image 5, this would mean the faces can be said to not include frontal-faces at all – in other words, the number of true frontal faces would be 0 and the face not picked up would be entirely fair. Ofcourse, this implies the false positive rate for this image would be high. However, our decision of annotations included all side faces as well as frontal faces, thus resulting in a TPR of 0.67. On assessment on the percentage overlap method used towards TPR, it was noted that other approaches existed, such as a comparison of medians. Additionally, a choice of sensitivity threshold would be varying depending on the spatial resolution of the image.

1.2 – Limitations Of TPR

The TPR on its own does not allow for much interpretation towards the accuracy of the classifier. A classifier producing a ratio of 10:90 true positives to false positives would yield the same TPR rate as a classifier with a ratio of 10:10 true to false positives. Hypothetically, we could to reduce the threshold parameters of our classifier as much as we possible; this would lead to an extremely higher acceptance rate. We could always find some threshold to give us a TPR of 1 – this consequently brings with it a detrimental following of false positives. Achieving a TPR of 1 provides us no assumptions that our classifier is precise.

1.3 – Formulating A Useful F1 Score

To calculated the F1 score we had to calculate the precision and recall (utilizing the method of calculating true positives explained before) where:

$$\text{Precision} = \frac{\text{True Positives}}{\text{Classifier Detections}} \qquad F = (p^2 + r^2)\frac{\text{Precision} * \text{Recall}}{P^2 * \text{Precision} + r^2 * \text{Recall}}$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{Annotations}}$$

In order to formulate an accurate F1 score, it should be taken into account how much bias we enforce onto the Precision or Recall. A traditional naive implementation, i.e. the harmonic mean, would weight both of these equally. The scaling of each co-efficient p and r positively correlates with how much importance we place onto the precision and recall respectively. The harmonic mean (p and r would both be 1) can be interpreted as us placing as much importance on both the accuracy (precision) and recall(TPR). This measure was used to classify the effectiveness of our classifier.

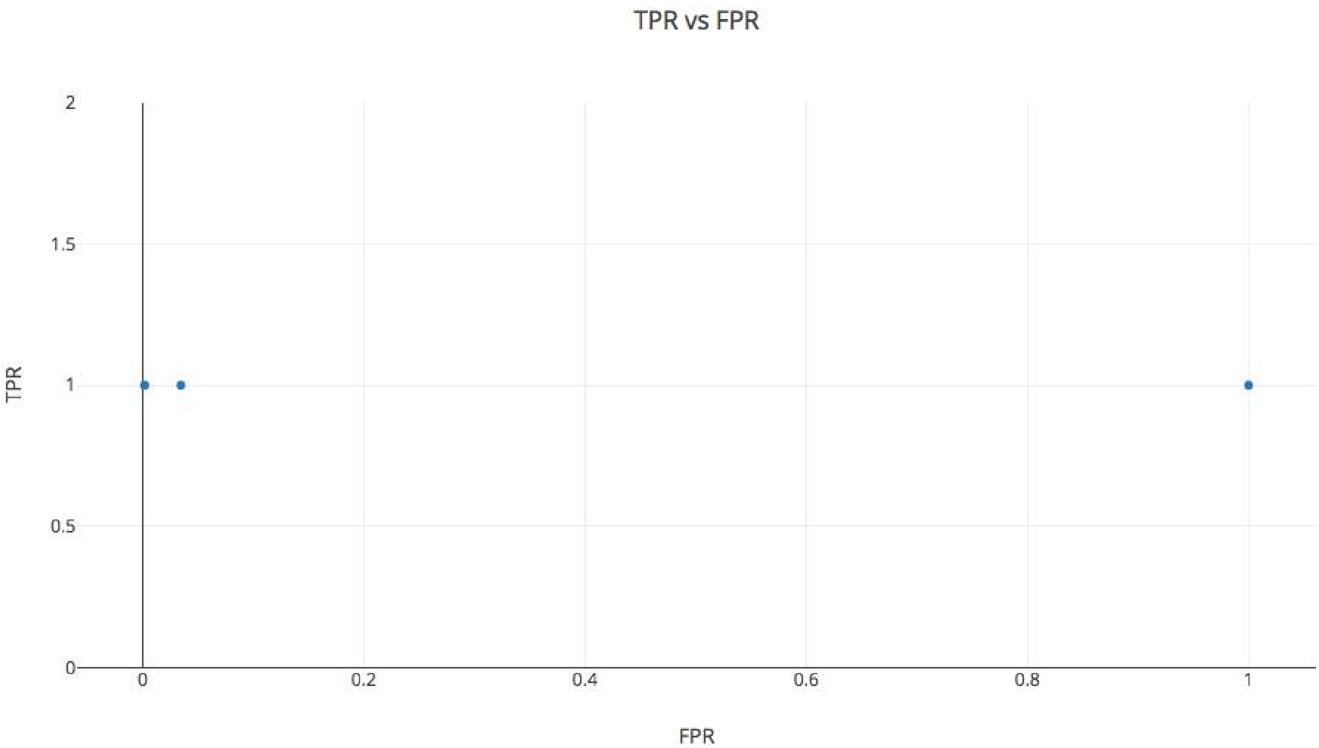
To produce a useful F1 representation, the relationship of weights applied to both the precision or recall should also be given a solid margin of acceptable difference in order to prevent inaccurate assessments of the classifier. This would prevent an extremely unbalanced pair precision and recall values from being unfairly scaled further.

As an example, given a classifier with high precision but very low recall, this would be extremely accurate however not at all useful even if we were to prefer precision. If we were to scale this further with a high precision factor p, the F1 would be a poor representative of the validity of the classifier. Conclusively, the measures we impose on our F1 formulation must account for the realistic bounds of the usefulness of our classifier regardless of our preference – i.e. how much are we are willing for our classifier to either pick up too many false positives, or too many false negatives, before its effectiveness renders it impractical as a tool.

Section 2 – Viola Jones Dartboard Detector Analysis

2.1 – TPR & FPR

During each stage of the attentional cascade more Haar-like features of reducing importance are accumulated into the resultant strong classifier. As we add more features the accuracy of our final classifier increases, seen by the reducing FPR at each stage (in Illustration 1). The top 10 features accumulated at stage 1 is sufficient in achieving a 100% TPR and thus each stage thereafter minimizes chances of false positives. The FPR is negatively correlated with the precision due to the FPR increasing as precision decreases. A low FPR after cascading training should be indicative of a high precision and (along with a high TPR) a high F1 score.



2.2 – Performance Analysis

During testing the TPR suffers especially in the case of occlusion in images 6, 11, 13, 15. As generated training data does not handle cases of occlusion this was expected. In the case of image 3, the image is illuminated from an angle this keeps one side of the image light and another dark, This is in contrast to generated training data were illumination of all pixels are scaled up or down. Haar-like features that are dependant on contrast in dark and light regions would not be able to pick up this dartboard. The classifier also fails in cases of extreme horizontal hilt in viewing angles.



Patterns, logos and complex shapes/objects with similar dark light regions were mispredicted as dartboards as a result of having similar haar-like features. These items however did not share geometric characteristics with a dart board, such as circular shape or lines meeting at a point. This motivated the need to integrate with the hough transform. The output statistics during training (FPR vs TPR plot) isn't representative of the accuracy of the classifier during testing.

Img	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Overall
TPR	1	1	1	0	1	1	0	1	1	1	0	0	1	0	1	0	0.625
Precision	1	1	0.16	0	1	0.5	0	1	0.25	0.5	0	0	0.33	0	0.28	0	0.376
F1	1	1	0.28	0	1	0.66	0	1	0.4	0.66	0	0	0.5	0	0.44	0	0.433

Section 3 – Hough Circle & Intersection Detector

3.1 – Examples Of Merits & Limitations

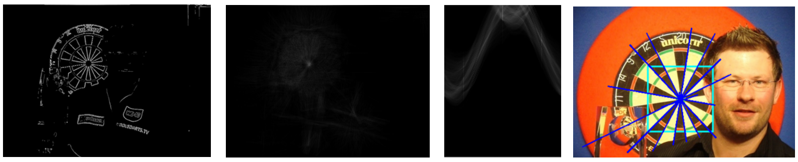


Image 4 - Merits

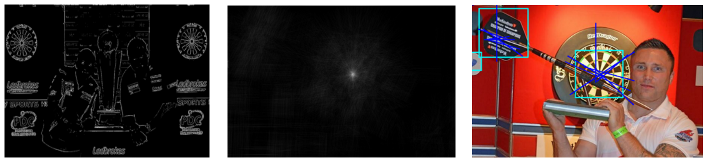


Image 13 - Limitations

Strong concentric circle detections were made as well as a clear central intersection for image 4. However for image 13, false intersections still persisted even for very strict line thresholds.

3.2 – Dartboard Detector Flowchart & Description

On analysis of the Viola-Jones detector, the largest concern was the amount of false negatives it produced as opposed to the false positives. Any false detections produced could easily be filtered through a second stage or even an alteration of threshold parameters.

Stage 1 – Hough Circles

We thus sought to create a dartboard detector using the Hough circle transform in order to provide a larger sample size to this filtering stage. We noted in Part 2 the weakness of Viola's Haar-like features to capture instances of primarily occlusion and less commonly deviation in angle. Hough circles' nature would let us detect partial circles as edge pixels as non-occluded sections of circles would contribute to the hough space.

We eliminated a large percentage of the noise in the image by the scaling pixel values up and thresholding voting pixels. A low hough threshold was set s.t. even smaller circles (i.e. further away dartboards) could be located. However, this unfortunately allowed noise to become more prominent in images where dart boards were closer. Typically the noise interfering was within the dartboard itself. This meant that many points within the actual dartboard would be identified as potential dartboard centres. To solve this, we found similar and concentric circles, merging these into a single circle with a center point of the highest Hough score.

Stage 2 – Splitting & Weighting Boxes

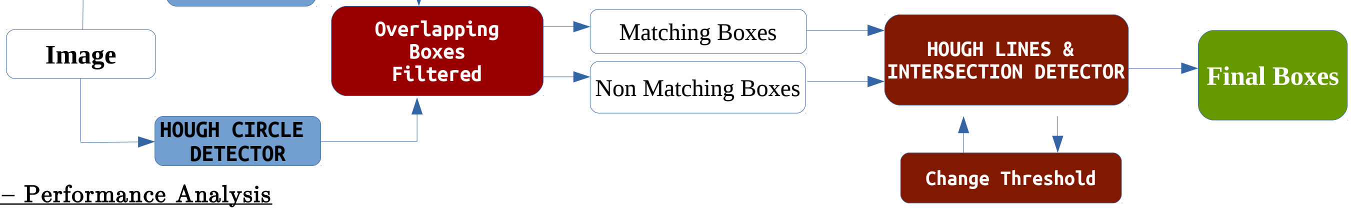
The output boxes from Viola-Jones and the circle detector were unioned after removing redundant boxes. It was found that for almost every image, identical boxes produced by Viola-Jones' detector and our Hough detector, correctly identified a true detection. We therefore chose to divide and weight these results into matching and non-matching bounding-boxes.

Stage 3 – Hough Lines & Intersections

For each box, the Hough line transform was applied to detect intersections occurring in the centre of the boxes produced by the previous stage. Similar lines were merged and then filtered in order to eliminate false intersections and reduce the initial large number of lines.

Lines not intersecting/inside the box area were removed. Lines were then elongated in order to increase probability of intersection. This was due to a much smaller number of lines at this stage, and an observation that many lines correctly corresponding to the true dartboard lines did not reach the centre. Common intersections found were then stored.

This whole stage was performed iteratively for each box over a range of thresholds, accumulating intersections. Finally, all accumulated intersections were filtered by their distance to box centres, and again thresholded by the minimum number of intersecting lines required. The resulting intersections, if any, confirmed a dartboard detection.



3.3 – Performance Analysis

IMAGE	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Average
Recall	1	1	1	1	1	1	1	1	0.5	1	1/3	0	1	1	1	0	**0.806**
Precision	1	1	1	1	1	0.5	1	1	1	1	1	0	1	0.5	1	0	**0.813**
F1	1	1	1	1	1	0.67	1	1	0.67	1	0.5	0	1	0.67	1	0	**0.782**

Merits:

- Our circle detector performed extremely well in improving the TPR of our classifier, successfully detecting occluded & angled dartboards.
- A main strength was the precision of detections, allowing us to improve significantly on the original Viola detector, which previously had an average F1 score of 0.462. A vast number of false positives were avoided due to filtering of redundant overlapping detections and additionally detection of intersections.
- The system of weighting identical detections between the circle and viola jones improved filtering through imposing a bias, being conservative in labelling the unsure cases of detections as true dartboard detections.
- Intersection detection was very successful. We solved issues of poor line samples through iterating through gradient magnitude and hough line thresholds. This accumulated a larger line representative to work with, balanced this out with increased thresholds in the later stages.
- Line elongation on a merged line sample allowed the detector compute a set of strong & sensible intersections.

Shortcomings:

- Intersection detection was highly dependent on a decent Hough line output. For a selection of images, the size of the Hough line sample were inadequate to identify intersections. The process of line merging would sometimes be a cause for this.
- The foundations of our true positive rate was only dependent on the Viola & Hough boxes.
- Tuning of intersection parameters, such as the threshold of distance for intersections to be considered 'the same', were difficult to balance between images – leading to better results in some, and worse in others.
- False intersections in boxes due to coincidental object alignment led to a 2 false positives overall.

Section 4 – Color Segmentation

4.1 – Rationale

We looked for an alternative, more accurate filtering stage to line and intesection detection. Line detections were generally inaccurate especially in distant dartboards as the dependencies of quality of the Hough line space were too dependant with parameters (such as line merge threshold, line intersection distance threshold) being very difficult to balance without achieving better precision in some images and worse in others. Therefore, non-intersecting boxes were now filtered using our implementation of color segmentation and (limited) scaling and sliding template matching to prune false positives.

4.12 – Implementation

We noticed alternating white and dark sectors as a continuous pattern within a dartboard - to extract this pattern we decided to perform color segmentation which clusters all pixels within an image between 2 groups. We first substituted pixels between the ranges denoting red and green in the HSV color space into black and white respectively. This would blend the inner ring of the dart board coloured in red and green with the colors of its associated sectors (as seen in ColFig2). We then performed a K-Means clustering in the RGB colors space (ColFig3). Afterwards, a binary thresholding followed by a thresholding operation within grid of pixels was performed (each grid typically 20-40pixels) that would set the value of a pixel to the average common pixel value of its neighbours in the grid (8-connectivity). Results given were typical to ColFig4 below.

The non-intersecting detections from before were clipped from the image, operated on by our color segmentation operation then template matched against a predefined template. The template matching essentially finds the hamming distance between 2 images, the first image the output of the color segmentation on a bounding box (our detection) and the second a default template (similar to ColFig4), we resize the template as necessary to fit the size of the bounding box before template matching.

ColFig 1

ColFig 2

ColFig 3

ColFig 4

Matching uses a very primitive sliding and scaling approach (unlike in viola which uses integral images) where we translate the image in the x and y by a given offset for different resized images and compute the hamming distance for each. The maximum hamming distance is considered in calculating its final match percentage. If the image is above a threshold if it not a false positive but a true dartboard. Since the template is resized for each sliding the aspect ratio of the template is changed (its width reduces) which allows it good matching with tilted dartboards in images.

4.2 – Merits of approach

The example images (before and after) exhibit the significant reduction in false positives after including this step.

4.3 - Merits vs Shortcomings

IMAGE	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Average
TPR	1	1	0	1	1	1	1	1	1	1	0	0	1	1	0.5	0	0.720
Precision	1	1	0	0.5	1	1	1	1	0.3	1	0	0	1	1	1	0	0.675
F1	1	1	0	0.67	1	1	1	1	0.5	1	0	0	1	1	0.67	0	0.698

Merits:

- Notice that the logo on the persons shirt (in the first pair of images) and object on the wall (in the second pair) were accurately filtered out. Violas haar-features evaluate a segment of the dartboard which typically would lead to these false positives, color segmentation and template matching performs a match against the general pattern of a dartboard. As opposed to viola, it is unaffected by matches between a segments/portions of the template and the image.
- Detecting by color is a more accurate comparison (comparatively to geometric shape comparison). Color segmentation resulted in a unique haar-like feature specific to dartboards, which detected them accurately.

Shortcomings:

- Still unable to detect certain dartboards in images viewed at a horizontal tilt. This effect is due to viola or hough circles being unable to detect these dartboards and thus never being accumulated into our final detections.
- Ideally we would add an additional detection step to enlarge our (to be filtered) sample size. The precision fell short comparatively against the intersection detection method resulting in an lower higher F1 score.
- The matches produced against boxes would occassionally be too high - even if the background were entirely black, about 50% of the template is black, meaning the match would be 50% even for a complete black detection. This would be problematic.