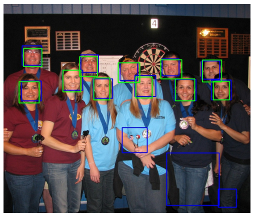
The Dartboard Challenge

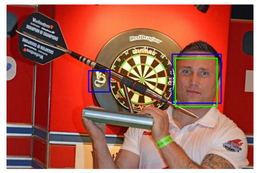
Callum Pearce cp15571, Ahmer Butt ab15015

1 The Viola-Jones Object Detector



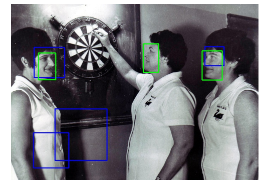
(a) dart5.jpg

(b) dart14.jpg



(c) dart13.jpg

(d) dart4.jpg



(e) dart15.jpg

Figure 1: Results of using the given frontal face AdaBoost classifier with Viola-Jones for detecting faces. Detected faces are shown in blue. Ground truths are shown in green. TPR for dart5 $= \frac{11}{11} = 1$, dart15 $= \frac{1}{1+2} = \frac{1}{3}$. Precision for dart5 $= \frac{11}{11+3} = \frac{11}{14}$, for dart15 $= \frac{1}{1+3} = \frac{1}{4}$. F_1-score for dart5 $= 0.88$, for dart15 $= 0.2857$

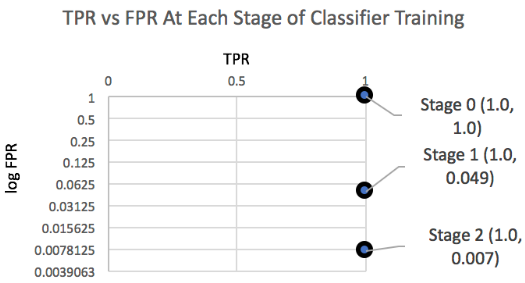
In Figure 1, the images *dart4*, *dart5*, *dart13*, *dart14* and *dart15* have been annotated with results of running the Viola-Jones Object Detector trained with the provided frontal face strong classifier. The results of object detection are illustrated by the bounding boxes drawn in blue. Ground truths are shown in green. The True Positive Rates, or *TPRs*, are included for *dart5* and *dart15*. There are a number of practical difficulties encountered when calculating the TPR. Problems arise because ground truths are not actually boxes, but should be precise outlines of a face. One could draw many different rectangles over a face and it may still be considered accurate. To compensate for this, the threshold of overlapping area between detected and ground truth bounding boxes for a true positive to be counted is reduced. But, having to choose a threshold adds human subjectivity. Consider the right-most face in image *dart15*. The region of the face that the detector picks up is small and misses out key features of the face, such as the mouth. In this report, this argument has been followed and so an overlap threshold such that this is not classified as a successful detection has been chosen.

It is always possible to achieve a TPR of 100% on any detection task. If a detector simply classifies every output as a positive, then there cannot be false negatives. Thus, a TPR of 100% is achieved. For the Viola-Jones detector discussed, this equates to classifying every possible Haar feature in every possible sliding window as a detected face. Therefore, one must consider more than just the TPR. Usually, the TPR is considered alongside the *precision*. This is the fraction of relevant positives among the detected positives. Whereas the TPR measures how *complete* the results are, precision measures how *correct* the results are. If a detector maximises TPR by detecting everything as positive, its precision will be very low. In a good system, precision decreases as the number of positives returned, or the TPR, increases. The F_1-score measures the precision/TPR trade-off by taking a harmonic mean. F_1-scores for *dart5* and *dart15* are included in Figure 1. In summary, F_1-scores are calculated given ground truth and a test run on any given test image as follows.

Algorithm 1 Calculating F_1-score for a test run.

procedure F1SCOREFORTESTRUN( )
  boxes ← coords of blue boxes and corresponding green boxes. if there is no corresponding box then add a corresponding box with zero coordinates
  T ← overlap threshold for detection
  for all pair of corresponding boxes in *boxes* **do**
    olap ← overlapping / ground truth area
    if *olap* == 0.0 **then** *false_negs++*
    else if *olap* == ∞ **then** *false_poss++*
    else if *olap* >= T **then** *true_poss++*
    else *false_negs++; false_poss++*
    end if
  end for
  $tpr \leftarrow true\_poss/(true\_poss + false\_negs)$
  $ppv \leftarrow true\_poss/(true\_poss + false\_poss)$
  $f_1 \leftarrow 2 * ppv * tpr/(ppv + tpr)$
  return f_1
end procedure

2 Building and Testing your own Detector



(a) Scatter plot of logFPR vs TPR after each training stage. Switching the axis gives a ROC curve.

F1 Scores for all 16 dartboards						
Dartboard	True Positives	False Negatives	False Positives	Recall	Precision	F1 Score
0	1	0	1	1.000	0.500	0.667
1	1	0	0	1.000	1.000	1.000
2	1	0	4	1.000	0.200	0.333
3	1	0	1	1.000	0.500	0.667
4	0	1	1	0.000	0.000	0.000
5	0	1	1	0.000	0.000	0.000
6	0	1	2	0.000	0.000	0.000
7	0	1	2	0.000	0.000	0.000
8	2	0	2	1.000	0.500	0.667
9	1	0	1	1.000	0.500	0.667
10	1	2	7	0.333	0.125	0.182
11	1	0	0	1.000	1.000	1.000
12	1	0	0	1.000	1.000	1.000
13	1	0	3	1.000	0.250	0.400
14	2	0	8	1.000	0.200	0.333
15	1	0	0	1.000	1.000	1.000
Average	**0.88**	**0.38**	**2.06**	**0.71**	**0.42**	**0.49**

(b) F_1-scores for the 16 dartboard images.

Figure 2

The TPR and FPR of the Viola-Jones AdaBoost classifier after each stage of training on the provided training data are shown as a scatter plot in Figure 2a. The plot shows that after each stage of training, the TPR remains at 100%, whereas the FPR drops. At each stage, Haar features are selected to build a weak classifier. Then, the weak classifiers are weighted and combined to build a strong classifier. In the first stage, the Haar feature chosen is very primitive so many image regions will match it. Therefore, the number of false negatives will be zero. The number of true positives will be maximised. The TPR will be 100%. In the proceeding stages, more Haar features are added. However, features that will reduce the TPR (on the training set) are purposefully not selected. Otherwise, the probability of not detecting a dartboard that does actually exist will increase from zero. The effect on the FPR is as follows. Selecting new, less trivial, features whilst maintaining a TPR of 100% results in rejecting more true negatives. This means that the true negatives will increase and the false positives will reduce. Therefore, the FPR decreases at every stage, as illustrated.

The TPRs and F_1-scores of the trained Viola-Jones detector on all the test data are summarised in Figure 2b. Example outputs of running the trained Viola-Jones detector on some of the test images are shown in the following Figure.

(a) dart0.jpg. All true positives detected, so TPR=0. But, one false positive, so precision=0.5 and F_1-score= 0.67.

(b) dart1.jpg. There is one true positive only. Therefore, both precision and TPR are perfect and there is an F_1-score of 1.

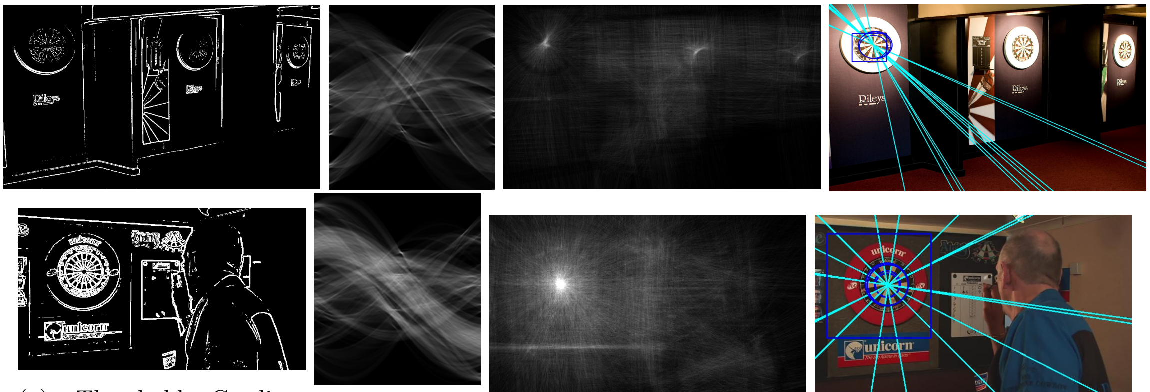
(c) dart4.jpg. There is no dartboard detected, and thus one false negative and zero true positives. So, TPR=precision=F_1-score=0 and FPR=1.

(d) dart10.jpg. There are lots of false positives. TPR is not affected but precision is very low. F_1-score is 0.18.

One can compare these scores with the scatter plot in Figure 2a. The scatter plot draws a ROC Curve (with axis switched) for the system. It shows that, on the training data, the TPR is always one and the FPR always falls at a fast rate. This is an almost perfect ROC curve. However, this is only for the training data. For the test data, the TPR is lower, at 0.7. This is because the training process overfits to the training data. An analogous pattern occurs when considering FPR. FPR measures the probability of falsely classifying as negative. The scatter plot shows that on the training data, the FPR is extremely low at 0.007. However, for the test data, there is an average of 0.375 false negatives per image. The perfect ROC curve is only possible due to overfitting to the training data. Therefore, one must appreciate this when using metrics derived from training data as a guide for predicting system performance on test data.

3 Integration with Shape Detectors

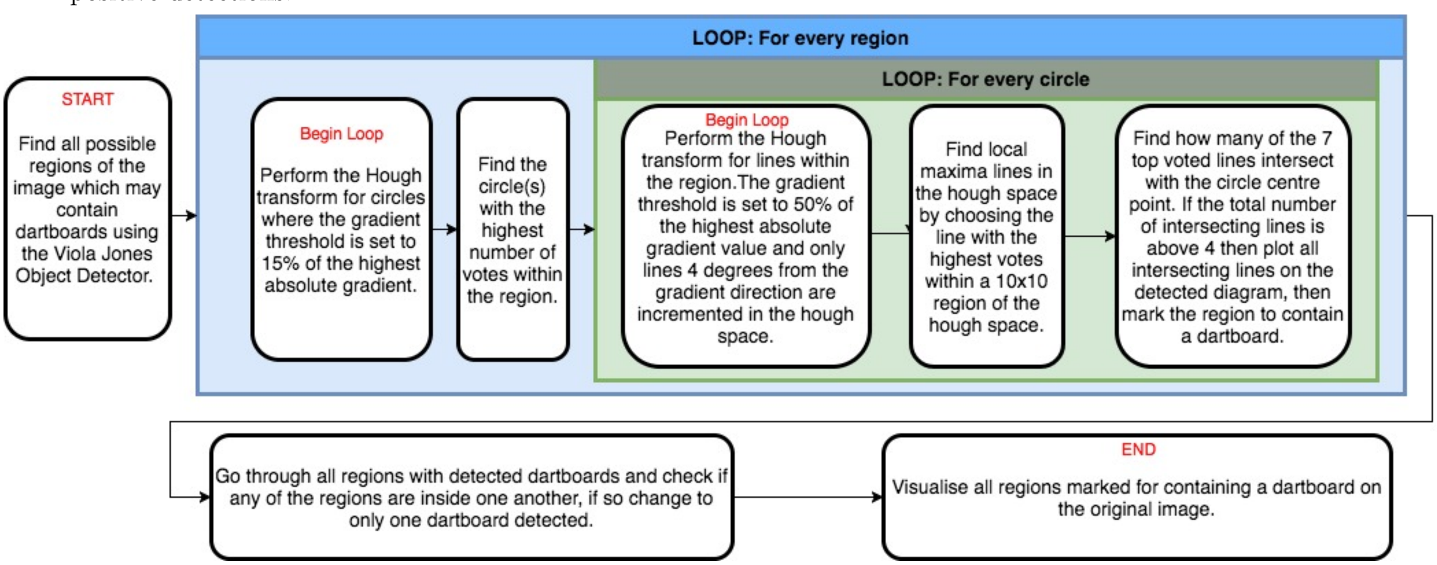
The Viola Jones detector discussed in the previous section is far from perfect, in this section shape detectors for both lines and circles are added to our detectors process. Using the Hough transform for lines and circles, key features of dartboards can be used as test cases to check whether a potential region of an image contains a dartboard. Below the results are displayed for the improved detector applied to test images 2 and 10, as well as the F1 score for all test images.



(a) Threshold Gradient Magnitude (b) Line Hough Space (c) Circle Hough Space (d) Detected Dartboards

DARTBOARD IMAGE	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	AVG
F1 SCORE	0	1	1	1	1	1	0	1	0.67	1	0.29	0	0	1	1	1	0.68

- The detector has a low chance of producing false positive detections due to its strict layered detection system which scrutinizes selective areas of the image. Including an initial Viola Jones pass followed by a Hough transform for circles, then a Hough transform for lines.
- A high probability in producing false negative detections is the detectors short coming. This is due to the success of the detector being completely dependent on the initial Viola Jones pass to detect all possible regions of the image which may contain a dartboard.
- Many circles/lines have an identical number of votes in the hough space, therefor choosing the highest n voted circles/lines is not always possible and multiple dartboards within a single region may be detected, resulting in false positive detections.



- Using Viola Jones as an initial estimate for finding the areas of possible dartboard locations heavily improved the detectors speed as the further detection techniques were more computationally intensive. Therefore optimizing areas of further detection reduced the runtime of the detector significantly.
- Circle detection is then applied to the image in order to classify the characteristic shape of a dartboard, rather then preforming line detection first as lines are far more frequent across the image, increasing the chance of producing false positive detections.
- Finally intersecting line detection on the center of the detected circle found the characteristic line pattern on a dartboard. This eliminated any circles within the Viola Jones region which is a common occurrence as most dartboards have number s and text around their perimeter including many circular shapes.

4 Improving your Detector

In this section a variety of techniques implemented to improve the accuracy of the dartboard detector are discussed. Each technique listed below directly improved one of the previous detectors shortcomings.

- **Altering the positive image generation and Viola Jones training parameters** in order to combat the high false negative rate produced by the detector. By generating more positive samples of dartboards (using 1000 compared to 500 in the previous section) with a higher max y value, the number of possible regions containing a dartboard increased. Increasing the max y value forced the positive image generation process to produce more dartboards with a elliptical shape rather then a perfect circle, this improved the accuracy in detection within test images such as dart12. In order to further increase the number of possible dartboard regions returned by the Viola Jones pass, the Max False Alarm Rate was altered to be a higher value (from 0.05 to 0.06). With more regions to scan the detector relied more on its sophisticated intersecting lines and circle detection process in order to filter which regions actually contain dartboards
- **Reducing the voting threshold until n lines/circles are found within the region** allows the detector to adapt to any individual image inputted into it. In the previous section for some images the detector would produce a high value of false negative images. This is partly due to the number of lines surpassing the voting threshold was far to low to confirm there was a dartboard in the region. If the voting threshold was reduced to combat this other images would have far too many lines surpassing the voting threshold, leading to a high false positive rate. By reducing the voting threshold until at least one circle was detected and 20 lines were detected, the dartboard detectors accuracy massively improved overall.
- **Finding the local maxima lines in a given area of the Hough space, then thresholding all other lines in the region** to reduce the the number of lines which have similar ρ and θ value. Implementing this local maxima and thresholding reduced the false positive rate produced by the detector, as there is a small case where a circle may be detected within a Viola Jones region even if a dartboard does not actually exist there. Then the line detector may find many intersecting lines at very small angles from one other, triggering the dartboard detector to confirm a dartboard is in the region. Therefor implementing a local maxima threshold ensures that all detected lines must be of a certain angle and distance from one another. This is a clear characteristic property of dartboards.

DARTBOARD IMAGE	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	AVG
F1 SCORE	1	1	1	1	1	0.67	1	1	0.57	0.67	0.5	0.67	1	1	0.8	1	**0.87**

- The optimizations made to the Viola Jones detector to increase the number of regions found which potentially contain a dartboard has led to a higher average F1 score. This is clear when comparing the table in the previous section with the one above, as there are now no zero F1 score values which previously resulted from a lack of detections from Viola Jones. Also the optimizations made to the Viola Jones detector improved its detection of more elliptical shaped dartboards, such as that seen in dart0 in 5.
- The maxima line optimization improved the F1 score for some images such as dart10 by ensuring that the most prominent lines detected in the region were evenly spaced. The effect is visualized in 6 where the local maxima is found for a 14x14 region within the dart0 line Hough space.
- Increasing the number of Viola Jones regions increased the number of false negative detections among the test images. This is visible in the detectors output image for dart8 in 5, as 2 false positive results are produced.

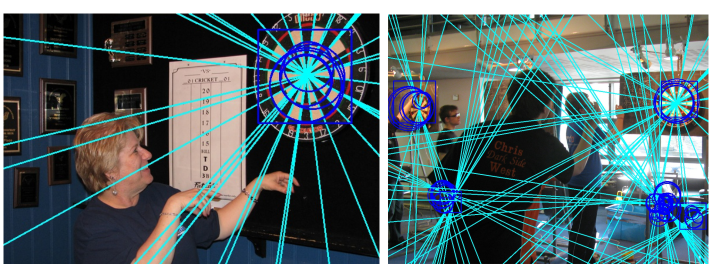


Figure 5: dart0 and dart8 with optimized detector results

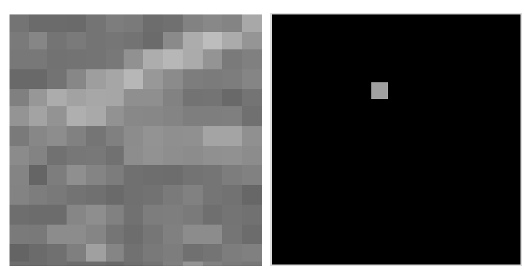


Figure 6: Local maxima line in dart0 Hough space