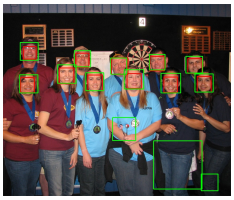
The Dartboard Challenge

By Tyler Ward and Andy Thorn (tw17530, at17556)

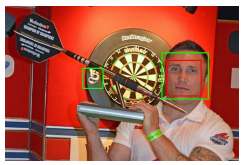
Subtask 1: The Viola-Jones Detector

Part a. Ground truth and visualisation

dart4.jpg

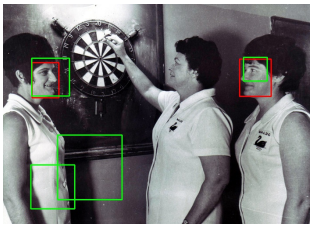


dart5.jpg



dart13.jpg

dart14.jpg



dart15.jpg

Figure 1.

Figure 1 shows five annotated images to show the accuracy of the Viola-Jones object detector. The red boxes are manually drawn and show the ground truth. The green boxes are the instances found by the detector.

Part b. IOU, TPR, F1-Score

Table 1.

Image	True Positive Rate (%)	F1-Score
Dart 0	0	0.00
Dart 1	0	0.00
Dart 2	0	0.00
Dart 3	0	0.00
Dart 4	100	1.00
Dart 5	100	0.88
Dart 6	0	0.00
Dart 7	0	0.00
Dart 8	0	0.00
Dart 9	100	0.40
Dart 10	0	0.00
Dart 11	100	1.00
Dart 12	0	0.00
Dart 13	100	0.67
Dart 14	100	0.50
Dart 15	50	0.33

Table 1 shows the true positive rate (TPR) for all of the test images given. It is a value calculated from the number of correctly identified front-facing faces and the total number of front-facing faces.

TPR is a good indicator of if your program is detecting faces in the image provided. However, it doesn't take into account if there are no faces in the picture. This makes it hard to distinguish between if faces not being detected or if there are no valid faces to detect. For example, the images Dart 1 and Dart 6 both have a 0% TPR, but Dart 1 has no valid faces in the image whereas Dart 6 does. Not knowing this could lead you to believe the detector is not detecting faces that are there in the case of Dart 1.

Due to the fact that most of the test images only had a few faces, the TPR was mostly either 100% or 0%. This is where we feel that using TPR as a gauge of the accuracy of the detector is a downfall. At a glance these numbers seem to show that the detector only works some of the time. Perhaps calculating the TPR for all of the images combined, or by using images containing more faces would give a better indication.

Because the true positive rate is a function of correctly detected faces over total valid faces, it does not penalise when the detector selects a region that is not a face. This means that a 100% true positive rate could be achieved with any image, if the program just selects every possible rectangle in the picture. This would ensure that every face is identified, but it would also output a lot falsely identified faces.

The final column in Table 1 show the F1-Score for each image. The average F1-Score over all of the test images was 0.30.

Subtask 2: Building and Testing your own Detector

Part a. Training Performance

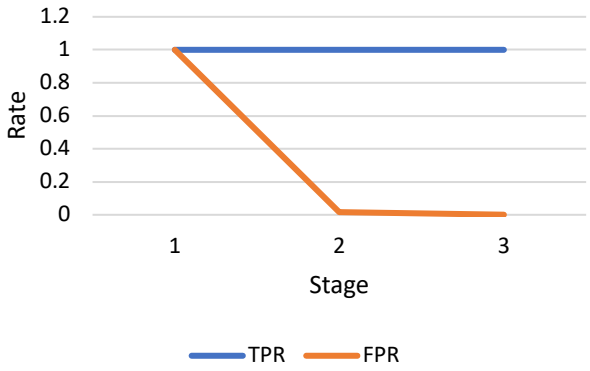
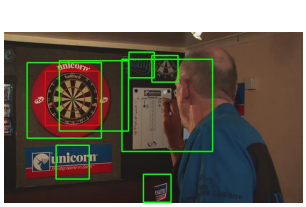


Figure 2.

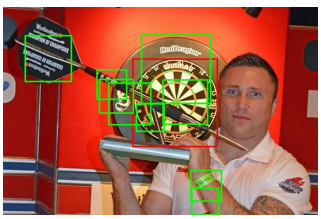
Figure 2 above plots the TPR vs FPR on the training data for the 3 different stages of training for the dartboard detector. As you can see from the graph, the untrained detector starts off with a TPR and FPR of 1; thus meaning that the classifier successfully identified all dart boards but also at the same time says everything is a dartboard. As training progressed through to stage 2, the false positive rate drops significantly. This signifies that the number of detections that aren't dartboards decreases while still maintaining the ability to correctly identify dartboards (TPR stays at 1). The move from stage 2 to 3 shows a lower FPR once again, however the improvement is not as significant. This is because the score is already very good, so it is hard for it to improve by the same amount.

Part b. Testing Performance



dart15.jpg

dart15.jpg



dart13.jpg

Figure 3.

The images in Figure 3 show the dartboard detections by our trained classifier (green boxes) and the ground truth (red boxes).

Table 2.

Image	True Positive Rate (%)	F1-Score
Dart 0	100	0.18
Dart 1	100	0.25
Dart 2	100	0.25
Dart 3	100	0.25
Dart 4	100	0.13
Dart 5	100	0.22
Dart 6	100	0.29
Dart 7	0	0.00
Dart 8	100	0.15
Dart 9	100	0.33
Dart 10	100	0.18
Dart 11	100	0.33
Dart 12	0	0.00
Dart 13	0	0.00
Dart 14	100	0.10
Dart 15	100	0.40

The average F1 score calculated from all 16 images was 0.19. This value is calculated by looking at both the precision and recall of the test; where a value closer to 1 represents near perfect precision (low false positives) and recall (low false negatives) and a value of 0 representing worse precision and recall. This indicator is better at correctly analysing a classifier's performance then just the TPR and FPR as it takes into account multiple factors like how precise and robust the classifier is. We also found that the plot generated from running the classifier on the training data to be unhelpful in predicting future test data. In training, the TPR remained at 100%, however upon using the classifier on real data it came to around 81.3%. This told us that the classifier was not able to successfully detect all dartboards 100% of the time like the plot suggested.

Subtask 3: Integration with Shape Detectors

Part a. Hough Details

Figure 4.

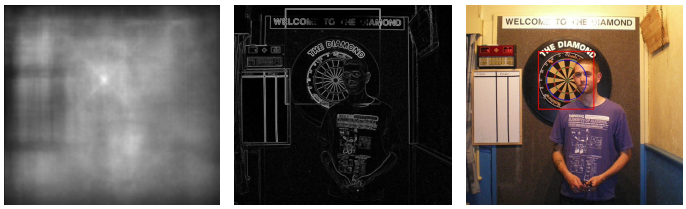


Figure 5.

Figures 4 and 5 display a 2D representation of the Hough space, the thresholded gradient magnitude and the result image respectively for two of the test images.

Part b. Evaluation

Table 3.

Image	True Positive Rate (%)	F1-Score	F1-Score compared to Viola Jones
Dart 0	100	1.00	+ 0.82
Dart 1	100	1.00	+ 0.75
Dart 2	100	1.00	+ 0.75
Dart 3	100	1.00	+ 0.75
Dart 4	0	0.00	- 0.13
Dart 5	100	1.00	+ 0.78
Dart 6	100	1.00	+ 0.71
Dart 7	0	0.00	0.00
Dart 8	50	0.67	+ 0.52
Dart 9	100	1.00	+ 0.67
Dart 10	33.3	0.50	+ 0.32
Dart 11	0	0.00	- 0.33
Dart 12	0	0.00	0.00
Dart 13	0	0.00	0.00
Dart 14	50	0.67	+ 0.57
Dart 15	100	1.00	+ 0.60

The average F1 score calculated from all 16 images was 0.62. This is a significant improvement when compared to the F1 score (0.19) of using just the Viola Jones Object Detector. One of the specific improvements we noticed when using a combination of both the Hough transform with the Viola Jones Object Detector was that the classifier was much more precise; meaning if it were to draw a box, it would be correct 93.8% of the time. The drawback of this increased precision was that the classifier was less robust (lower recall) than it was previously; meaning that it started to miss some dartboard detections which it would detect previously. These two factors are a trade-off and we chose to focus on precision rather than robustness.

Key merits

- Increased precision over the Viola Jones Object Detector
- Easily tuneable - can change the thresholds to adapt to different working environments
- Able to find frontal view dartboards even with obtrusions
- Works in low lighting conditions
- Does not get fooled between circular objects and dartboards

Shortcomings

- Bottlenecked by Viola Jones Object Detector - if the Viola Jones classifier did not detect a dartboard, this new classifier wouldn't be able to either as we use a combination of both the Hough transform and the Viola Jones classifier outputs to detect boards
- Lower recall then the Viola Jones Object Detector
- Could use the line pattern of the dartboard in conjunction with circle detection to aid dartboard detection
- Can't detect ellipses
- Can't detect more than one board per image

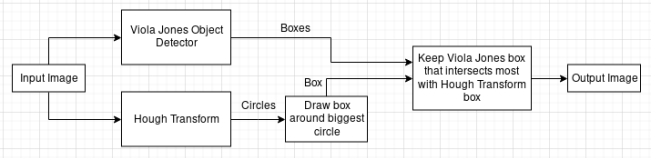


Figure 6.

Subtask 4: Improving your Detector

Although we did not end up adding any improvements to our solution for subtask 3, we discussed multiple improvements to further improve the detectors accuracy.

- Detecting ellipses
  Our detector currently only looks for circles in the image to identify dartboards and uses the minimum and maximum radius thresholds we set to detect where dartboards are. If the picture is taken at a slight angle the dartboard appears as an ellipse, to improve our detector we would first allow for our program to detect ellipses.

- Detecting multiple intersecting lines
  Next we would search for multiple lines that intersect at a middle point and combine this with our circle detection. If the midpoint of the circle matches up the line intersect, the selection would be classified as a dartboard. This would reduce the number of false positives the program outputs as detecting lines and circles provides a sort of "check" system.

- Detecting multiple dartboards
  Finally, our detector assumes that there is only one dartboard in a picture. If there are multiple visible boards in the picture, the largest one is selected. This problem could be worked around if we added line detection functionality as well as circle detection we could allow for the program to select multiple dartboards.

- Viola Jones Detector Improvement
  One of the major drawbacks of our detector was that the Viola-Jones detector we used wasn't robust enough, meaning that sometimes the dartboard was not detected even though our Hough transform detected the circle. Therefore, the use of a Viola-Jones detector with a higher recall rate (accepting lower precision) would significantly improve our detector.

Signatures

Tyler Ward (tw17530)
Contribution weight: 1

Andy Thorn (at17556)
Contribution weight: 1