IPCV-Coursework

Luca Gough

1 Shape Detection

1.1 Viola-Jones training performance

An Attentional Cascade Viola-Jones detector[2] was trained on a set of positive and negative sample images. This detector uses four Haar-like features which represent the sum of different pixel areas within the feature. These values can indicate the existence of features within an image. A thresholded weighted sum of the Haar-like features can be used to classify images with certain features. The weights can trained using the Adaboost Algorithm, which attempts to minimise the error of the classifier by adjusting the weights. Finally, a cascade utilises a chain of classifiers that use increasing numbers of Haar-like features, a negative result at any layer of the cascade results in immediate rejection. Whereas a positive results leads to processing in a further layer.
The log of the true positive rate and the false positive rate at each stage of training the cascade is shown in Figure 1, log scales were used as the FPR values are very small. At stage zero the classifier has a false acceptance ratio of 1:500, eliminating the most negative samples of all the stages. Stage one has a false acceptance ratio of 0.06:500 since a positive sub-window from the first classifier triggers the evaluation of the second classifier. The second layer is only rejecting a small number of negative samples compared to the first layer. Similarly in the third layer the false positive ratio is 0.008:500. At every stage the True Positive acceptance rate is 1 meaning all positive samples are accepted.

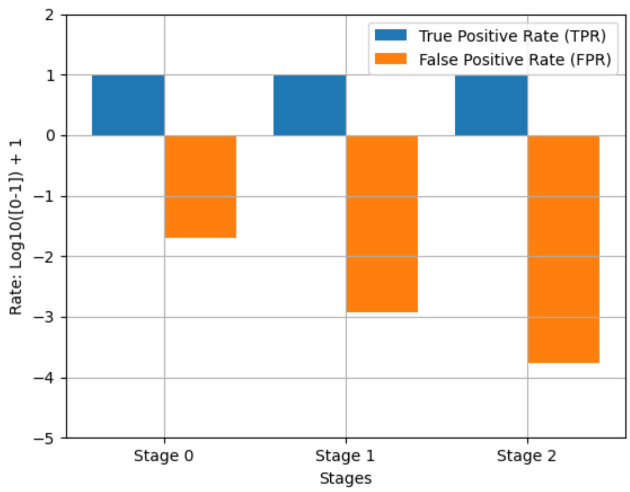


Figure 1: Log TPR and FPR of the 3-layer cascade

1.2 Viola-Jones testing performance

During testing the model was able to correctly identify 59% of the dartboards across 16 test images. Notably it was able to correctly identify the dartboard in Figure 2 image 4, which was partially occluded.

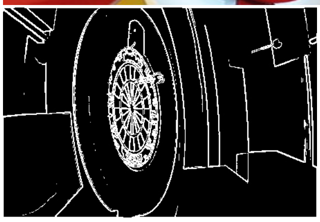
Image	TPR	F_1
0	1.0	0.67
1	1.0	1.00
2	1.0	0.67
3	1.0	0.40
4	1.0	1.00
5	1.0	0.50
6	0.0	0.00
7	0.0	0.00
8	0.5	0.20
9	1.0	1.00
10	0.0	0.00
11	0.0	0.00
12	0.0	0.00
13	0.0	0.00
14	1.0	0.36
15	1.0	1.00
Average	0.59	0.43

The F_1 score can be interpreted as the mean of the precision and recall of the model. The F_1 score averaged 0.43; this outcome is clearly evident in Figure 2. In image 14 the model correctly identifies both dartboards leading to a high recall score but it predicts seven false positives resulting in a low precision score. This tells us that training set of negative samples may not have been diverse enough as the model consistently finds false positives in the more complex images.

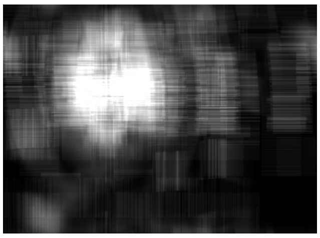
Figure 2: Top: Image 14, Left: Image 4, Right: Image 8. Red Box = Ground Truth, Blue box = Detection

1.3 Hough Transform Integration

In order to improve the overall performance of the detector Hough Transforms we're used to detect circles and ellipses. For each image a thresholded gradient image (Figure 3a and Figure 4a) was produced using horizontal and vertical Sobel filters.



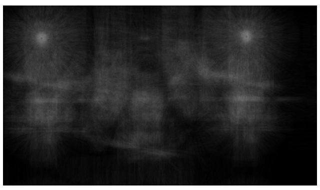
(a) Thresholded Gradient Magnitude



(b) Ellipse Hough Space

Figure 3: Image 12 Detection

(a) Thresholded Gradient Magnitude



(b) Circle Hough Space

Figure 4: Image 14 Detection

The advantages of detecting ellipses was that it allowed the detector to identify dartboards that we're at an angle to the camera, however this was very slow to compute so this was only used if the circular Hough space didn't contain any significant results. The drawbacks of the circle detector is that it found lots of false positives for example in Figure 4 the circular trophy is detected with high confidence.

1.4 Shape Detector Evaluation

- Increased True Positive rate implies the shape detectors are able to identify dartboards that the Viola Jones detector couldn't find.
- Increased F_1 score means that the detector produced less false positives and left less false negatives.
- Ellipse detector was extremely slow taking 2 minutes for a single image and used large amount of memory.
- The Ellipse detector requires the target ellipse to be closely oriented with the x and y axis, as adding an angle to the Hough space increases its size exponentially.
- Thresholding values for significant ellipses/circles most likely overfits the model to this test data.

			Viola Jones Δ	
Image	TPR	F_1	TPR	F_1
0	1.0	1.00	0.00	+0.33
1	1.0	1.00	0.00	0.00
2	1.0	1.00	0.00	+0.33
3	1.0	1.00	0.00	+0.60
4	1.0	1.00	0.00	0.00
5	1.0	0.67	0.00	+0.17
6	0.0	0.00	0.00	0.00
7	1.0	1.00	+1.00	+1.00
8	0.5	0.50	0.00	+0.30
9	1.0	1.00	0.00	0.00
10	0.67	0.80	+0.67	+0.80
11	0.0	0.00	0.00	0.00
12	1.0	1.00	+1.00	+1.00
13	1.0	1.00	+1.00	+1.00
14	1.0	0.67	0.00	+0.31
15	1.0	0.67	0.00	-0.33
Average	0.82	0.77	+0.23	+0.34

Figure 5: Hough Circle and Ellipse Performance

1.5 Detection Pipeline

- The first way a detection can arise is if the IoU score between a Viola-Jones detection and a Circle Detection is greater than some threshold, this reduces false positives from Viola-Jones.
- If the confidence, Hough space count, of a circle detection is high enough a dartboard is detected.
- If there are circle detections but of low confidence, this implies the existence of a curved surface which may be better identified with an ellipse.

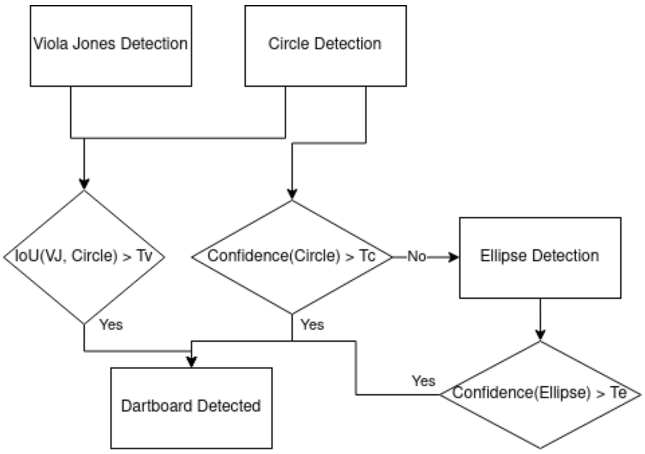


Figure 6: Dartboard Detector Pipeline

1.6 Detector Improvements

The fact that dartboards contain a circle of alternating coloured triangles that point towards the centre makes a very identify-able pattern inside each dartboard. In order to recognise these triangles first a set of contours we're acquired from a thresholded binary gray-scale image. OpenCV's 'findContours()' uses a border following method by scanning through the image tracing the border of each region of different colours to establish a hierarchy of contours within the image [1].
The closed contours of the image will have many different edges as the boundaries in the image will not be perfectly straight. OpenCV's approxPolyDP uses the Ramer–Douglas–Peucker algorithm to simplify a polyline into a line with fewer points. This is achieved by recursively subdividing the initial line, keeping the first points that are less than a given approximation value. This is generally some proportion of the length of the original contour. The resulting polylines were filtered to only include those with 3 sides.
This resulted in a list of triangles, these we're then filtered to only include those that contained an angle of 18°(A dartboard contains 20 triangular segments, $\frac{20}{360} = 18$). Finally a basic clustering technique is used to identify areas containing 3 or more of these triangles. These areas are used as the bounding box for the dartboard detection.

- The previous two components we're unable to detect dartboards with large occlusions, Figure 8 highlights how the detector only needs a few visible triangles
- Figure 8 Also shows how the detector is unable to identify dartboards that are far away due to the small geometry

			Viola Jones + Hough Δ	
Image	TPR	F_1	TPR	F_1
0	1.0	1.00	0.00	0.00
1	1.0	1.00	0.00	0.00
2	1.0	1.00	0.00	0.00
3	1.0	1.00	0.00	0.00
4	1.0	1.00	0.00	0.00
5	1.0	0.67	0.00	0.00
6	1.0	1.00	+1.00	+1.00
7	1.0	1.00	0.00	0.00
8	0.5	0.50	0.00	0.00
9	1.0	0.67	0.00	-0.33
10	0.67	0.80	0.00	0.00
11	1.0	1.00	+1.00	+1.00
12	1.0	1.00	0.00	0.00
13	1.0	1.00	0.00	0.00
14	1.0	0.57	0.00	+0.10
15	1.0	0.67	0.00	0.00
Average	0.95	0.87	+0.125	+0.11

Figure 7: Hough Circle and Ellipse Performance

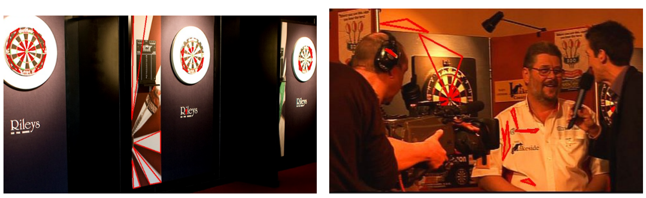


Figure 8: Contours filtered to 3 sides after polygon approximation

References

[1] Satoshi Suzuki and KeiichiA be. Topological structural analysis of digitized binary images by border following. *Computer Vision, Graphics, and Image Processing*, 30(1):32–46, 1985.

[2] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, volume 1, pages I–I, 2001.