

Image Processing and Computer Vision
COMS30031

Kai Hulme
kh16747
University of Bristol
26th July 2021

1 Face Detection with Viola-Jones

1.1 Ground Truth and Detection Visualisation



Figure 1: Results of Viola-Jones face detection on images *dart4.jpg*, *dart5.jpg*, *dart13.jpg*, *dart14.jpg* and *dart15.jpg* (from top-left to bottom-right), with ground truth bounding boxes in red and detections in green. Performance on the top row of images is perfect, with no false positives (FP) or false negatives (FN). The lower-left image has detected both faces but has mistakenly detecting the trophy as a face. The lower-right image has detected the face of the most forward facing woman but has missed the two on the right who are at a greater angle to the camera, indicating it struggles to detect less straight-on faces - likely due to less facial features being visible.

1.2 IoU, Precision, Recall and F_1 -Score

$$IoU_{a,b} = \frac{A_{a \cap b}}{A_{a \cup b}} \quad (1) \quad precision = \frac{TP}{TP + FP} \quad (2) \quad recall = \frac{TP}{TP + FN} \quad (3) \quad F_1 = \frac{2 \cdot precision \cdot recall}{precision + recall} \quad (4)$$

When evaluating an object detector we are interested in its ability to correctly detect what we are looking for, e.g. faces. A correct detection is a true positive (TP), a false detection is a false positive (FP) and a missed detection is a false negative (FN). To check if a detection is correct, the detected bounding box is compared to the ground truth bounding using the Intersection over Union (IoU) (1). Non-intersecting boxes give an IoU of 0 (as $A_{a \cap b} = 0$) and 1 for perfectly intersecting boxes (as $A_{a \cap b} = A_{a \cup b}$); a detection is correct if the IoU of the detection and ground truth is greater than some threshold value.

Recall (3), also called the true positive rate, measures the proportion of positive detections out of all the positives, effectively measuring the detectors ability to detect. However, recall alone should not determine the performance of a detector. A naïve detector could return every possible bounding box which would include the ground truths, but this would not be of use as there are so many FPs. Precision (2) measures the proportion of correct detections from all of the detections. A precise detector will still aim to find TPs, but reduce FPs. F_1 -score (4) is the harmonic mean of precision and recall, meaning a detector with a high F_1 will have both high precision and recall which should result in accurate detections with few mistakes. Table 1 shows the precision, recall and F_1 -score of the face detection results from the Viola-Jones cascade detector on each of the test images.

Recall (or TPR) cannot be measured accurately when there is nothing to detect. Looking at (3), if there are no true positives or detections precision is undefined. For example: with $TP = 0$ and $FN = 0$, $recall = \frac{0}{0+0}$. For this reason images with no faces in Table 1 have *n/a* results.

Image	Precision	Recall	F_1
<i>dart0</i>	1.00	1.00	1.00
<i>dart1</i>	<i>n/a</i>	<i>n/a</i>	<i>n/a</i>
<i>dart2</i>	<i>n/a</i>	<i>n/a</i>	<i>n/a</i>
<i>dart3</i>	<i>n/a</i>	<i>n/a</i>	<i>n/a</i>
<i>dart4</i>	1.00	1.00	1.00
<i>dart5</i>	0.85	1.00	0.92
<i>dart6</i>	0.00	0.00	0.00
<i>dart7</i>	1.00	1.00	1.00
<i>dart8</i>	0.00	0.00	0.00
<i>dart9</i>	1.00	1.00	1.00
<i>dart10</i>	<i>n/a</i>	<i>n/a</i>	<i>n/a</i>
<i>dart11</i>	1.00	1.00	1.00
<i>dart12</i>	<i>n/a</i>	<i>n/a</i>	<i>n/a</i>
<i>dart13</i>	1.00	1.00	1.00
<i>dart14</i>	0.67	1.00	0.80
<i>dart15</i>	0.33	0.33	0.33

Table 1: Table showing detection performance of Viola-Jones for each test image. *n/a* entries indicate no faces in test image which would result in undefined precision, recall and F_1 metrics.

2 Dartboard Detection with Viola-Jones

2.1 Training Performance

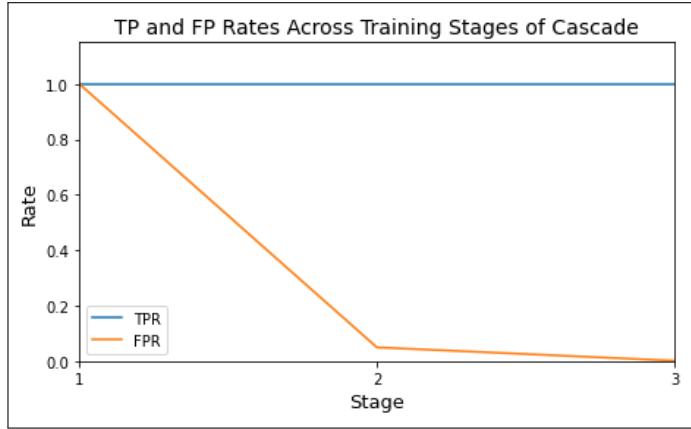


Figure 2: True positive rate (TPR) and false positive rate (FPR) throughout 3 stages of training Viola-Jones cascade for dartboard detection. The Viola Jones algorithm is trained in a number of stages. At each stage a new layer will be added to the detector in the form of a cascade. The training results indicate the cascade favours maximising detections first, giving a TPR close to 1.00 at the expense of the FPR being 1.00 As training progresses the cascade becomes more precise with the FPR dropping close to 0.00 by stage 3. The final cascade will have a high F_1 close to 1.00 when tested on the training images.

2.2 Testing Performance



Figure 3: Results of Viola-Jones dartboard detection on images *dart2.jpg*, *dart11.jpg*, *dart14.jpg* (from left to right), with ground truth bounding boxes in red and detections in green.

Figure 3 shows the detections made by the Viola-Jones cascade trained to detect dartboards. Ground truth bounding boxes are shown in red and the cascade's predictions are shown in red. I used a threshold of 0.5 for the IoU between the boxes to discern TPs from FPs.

Looking at the first image we can see the Viola-Jones method has detected the target dartboard well, but there are a large number of FPs resulting in a recall of 1.00, but a precision of only 0.14 giving an F_1 of 0.25. The second image shows a missed detection (FN), likely caused by the dartboard being obstructed by camera, there is also a FP in the upper-right corner giving an F_1 of 0.0. The image on the right shows the system is able to detect multiple dartboards. However, not only are there are FP detections, but the dartboards have been detected multiple times (multiple detections are FPs), meaning while recall is 1.00, precision is 0.25.

Overall it seems the Viola-Jones method works well for non-obstructed dartboards. However the system has many FPs and its precision and consequently its F_1 suffers from this. Testing gives a high recall of 0.90, but a low precision and F_1 of 0.52 and 0.61 respectively. When improving this system I aim to increase the precision whilst trying to retain its high recall.

The testing results in Table 2 differ from those seen during training in Figure 2.1. This indicates the Viola-Jones method has some degree of overfitting to the training data and so performance suffers when tested on unseen data. This is likely due to the nature of the training data differing from the test data; the positive samples in training were all augmentations of the same dartboard; where the testing data is more varied with different board types, lighting conditions, obstructions which lead to FNs, as well as other items in the image which lead to FPs.

Image	Precision	Recall	F_1
<i>dart0</i>	1.00	1.00	1.00
<i>dart1</i>	0.50	1.00	0.67
<i>dart2</i>	0.14	1.00	0.25
<i>dart3</i>	0.33	1.00	0.50
<i>dart4</i>	1.00	1.00	1.00
<i>dart5</i>	0.20	1.00	0.33
<i>dart6</i>	1.00	1.00	1.00
<i>dart7</i>	0.50	1.00	0.67
<i>dart8</i>	0.20	0.50	0.29
<i>dart9</i>	0.25	1.00	0.40
<i>dart10</i>	0.43	1.00	0.60
<i>dart11</i>	0.00	0.00	0.00
<i>dart12</i>	1.00	1.00	1.00
<i>dart13</i>	0.50	1.00	0.67
<i>dart14</i>	0.25	1.00	0.40
<i>dart15</i>	1.00	1.00	1.00
avg.	0.52	0.90	0.61

Table 2: Dartboard detection performance of Viola-Jones method for each image and overall

3 Integration with Shape Detectors

3.1 Hough Details

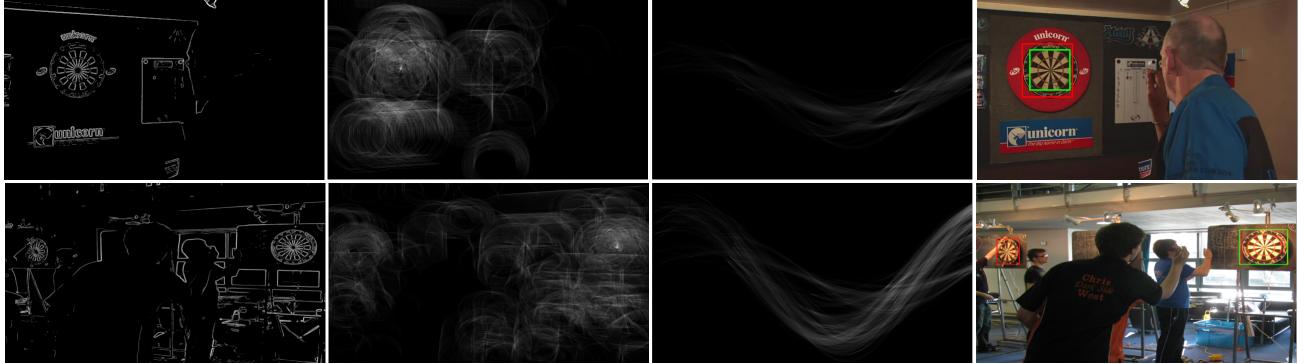


Figure 4: Thresholded gradient magnitude from Sobel edge detection, Hough (circles) space at radius close to ground truth, Hough (lines) space and detections for *dart2* (top row) and *dart8* (bottom row). Images have been resized.

3.2 Evaluation

- Looking at Figure 4 my ensemble method was able to detect boards where the Viola-Jones method missed, whilst greatly reducing the number of FPs.
- Table 3 shows a precision increase of 0.18 due to far less FPs. The detection rate suffered but the Viola-Jones method had such a high recall due to its low precision; if you say everything is a dartboard you'll not miss any.
- Overall the system had an increase in F_1 score whilst being far more precise than the Viola-Jones method alone. Further improvements will aim to increase the recall.

3.3 Detection Pipeline

- As my aim was to create a more precise system I decided to use an ensemble of detectors and require two methods to agree on a detection before accepting it.
- The 3 methods are the original Viola-Jones cascade along with circle and line detection via Hough transforms.
- Dartboards are circular with lines separating scoring zones, so a circle with enough lines passing through must be a dartboard. Checking if a Viola-Jones detection has either a circle centre or enough lines passing through would also result in a detection.
- Additional checks were added to ensure each dartboard was only detected once.

Image	Precision	Recall	F_1
<i>dart0</i>	1.00 (+0.00)	1.00 (+0.00)	1.00 (+0.00)
<i>dart1</i>	1.00 (+0.05)	1.00 (+0.00)	1.00 (+0.33)
<i>dart2</i>	1.00 (+0.86)	1.00 (+0.00)	1.00 (+0.75)
<i>dart3</i>	0.00 (-0.33)	0.00 (-1.00)	0.00 (-0.50)
<i>dart4</i>	1.00 (+0.00)	1.00 (+0.00)	1.00 (+0.00)
<i>dart5</i>	0.00 (-0.20)	0.00 (-1.00)	0.00 (-0.33)
<i>dart6</i>	0.00 (-1.00)	0.00 (-1.00)	0.00 (-1.00)
<i>dart7</i>	1.00 (+0.50)	1.00 (+0.00)	1.00 (+0.33)
<i>dart8</i>	1.00 (+0.80)	0.50 (+0.00)	0.67 (+0.38)
<i>dart9</i>	0.50 (+0.25)	1.00 (+0.00)	0.67 (+0.27)
<i>dart10</i>	1.00 (+0.57)	0.67 (-0.33)	0.80 (+0.20)
<i>dart11</i>	0.00 (+0.00)	0.00 (+0.00)	0.00 (+0.00)
<i>dart12</i>	1.00 (+0.00)	1.00 (+0.00)	1.00 (+0.00)
<i>dart13</i>	1.00 (+0.50)	1.00 (+0.00)	1.00 (+0.33)
<i>dart14</i>	0.67 (+0.42)	1.00 (+0.00)	0.80 (+0.40)
<i>dart15</i>	1.00 (+0.00)	1.00 (+0.00)	1.00 (+0.00)
avg.	0.70 (+0.18)	0.70 (-0.20)	0.70 (+0.09)

Table 3: Dartboard detection performance of ensemble method for each image and overall with improvements over Viola-Jones method shown for each entry.

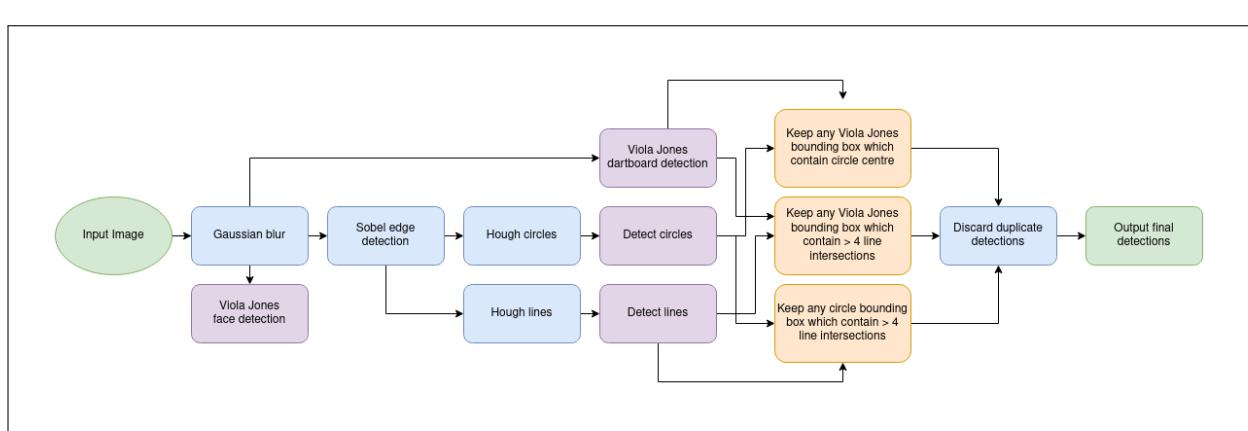


Figure 5: Flow diagram depicting ensemble method of Viola-Jones with Hough transforms for dartboard detection.

4 Improving my Detector

4.1 Idea and Visualisation

- The first improvement I made was to the Hough transforms; using gradient angles obtained during Sobel edge detection I reduced the range of angles used to increment the accumulator for both Hough lines and circles. This reduced the amount of noise within the Hough space caused by intersections of points of non-interest accumulating to high values which result in FPs. Figure 6 shows this reduction in noise resulting in an accurate circle detection. This method greatly increased precision but I included the results with the previous section in Table 6.
- The second improvement was to cluster the colours of the image using KMeans. When inspecting some of the images I noticed many FP detected circles; Figure 6 shows this in *dart3.jpg*. Using color segmentation I hoped problem areas would be segmented together due to similarities in colour, removing edges causing the FPs. Looking at *dart3.jpg*, yellow curved border was removed and the scoring zones of the dartboard have more contrast allowing for my ensemble method to detect it due to many detected lines intersecting the Viola-Jones detection bounding box.



Figure 6: Top row *dart2.jpg*: Hough circles space (left), reduced Hough space (middle), TP detected circle (right). Bottom row *dart3.jpg*: FP detected circle (left), KMeans colour segmented (middle), TP detected dartboard (right).

4.2 Evaluation

- I aimed to increase the precision of the Viola-Jones method. Final results in Table 4 show a precision increase 0.22 over the Viola-Jones method whilst retaining high recall, resulting in an F_1 of 0.8; an increase of 0.19. Additionally average IoU of detections increased by 22% over Viola-Jones.

- Circle detection works well but dartboards are sometimes not straight-on and so are often not circular. Ellipse detection would be a better option for off-centre boards.
- Lines detection works excellently, but edges are prevalent throughout the image. Using line detection as a method of validating other methods was a good decision.
- The addition of another shape detector such as triangles would improve recall further whilst retaining precision if added to current ensemble.
- Overall my Hough transforms work well and the improvements discussed in 4.1 increased precision greatly.
- KMeans colour segmentation reduced the number of edges in the image reducing FPs.
- Performance of the system is largely down to parameter selection. Searching the parameter space was limited by the speed of Python, however I have minimised processing time via vectorisation of complex functions.
- Greater performance is likely possible but I am pleased with my current implementation and more parameter optimisation would not add further merit to my project.

Image	Precision	Recall	F_1
<i>dart0</i>	1.00 (+0.00)	1.00 (+0.00)	1.00 (+0.00)
<i>dart1</i>	1.00 (+0.00)	1.00 (+0.00)	1.00 (+0.00)
<i>dart2</i>	0.50 (-0.50)	1.00 (+0.00)	0.67 (-0.33)
<i>dart3</i>	1.00 (+1.00)	1.00 (+1.00)	1.00 (+1.00)
<i>dart4</i>	1.00 (+0.00)	1.00 (+0.00)	1.00 (+0.00)
<i>dart5</i>	0.33 (+0.33)	1.00 (+1.00)	0.50 (+0.50)
<i>dart6</i>	1.00 (+0.00)	1.00 (+1.00)	1.00 (+1.00)
<i>dart7</i>	1.00 (+0.00)	1.00 (+0.00)	1.00 (+0.00)
<i>dart8</i>	0.50 (-0.50)	0.50 (+0.00)	0.50 (-0.17)
<i>dart9</i>	1.00 (+0.50)	1.00 (+0.00)	1.00 (+0.33)
<i>dart10</i>	0.67 (-0.33)	0.67 (+0.00)	0.67 (-0.13)
<i>dart11</i>	0.00 (+0.00)	0.00 (+0.00)	0.00 (+0.00)
<i>dart12</i>	1.00 (+0.00)	1.00 (+0.00)	1.00 (+0.00)
<i>dart13</i>	0.50 (-0.50)	1.00 (+0.00)	0.67 (-0.33)
<i>dart14</i>	0.33 (-0.34)	1.00 (+0.00)	0.50 (-0.30)
<i>dart15</i>	1.00 (+0.00)	1.00 (+0.00)	1.00 (+0.00)
avg.	0.74 (+0.04)	0.89 (+0.19)	0.80 (+0.10)

Table 4: Dartboard detection performance of final ensemble method with KMeans colour segmentation for each image and overall with improvements over ensemble method shown for each entry.