# The Dartboard Challenge

Image Processing and Computer Vision

**Kipp McAdam-Freud and Stan McCarthy**

# 1 Viola-Jones Object Detection

To quantify the accuracy of our Viola-Jones detector, we first define three terms:

- **True Positive** (TP): A face in our image which has been accurately detected by our program.
- **False Positive** (FP): An area which has been incorrectly flagged as containing a face by our detector.
- **False Negative** (FN): A face in our image which has not been detected by our program.

To calculate these values for each image, a program was written which compared the frames generated by our Viola-Jones detector to “ground truth” rectangles; rectangles created by the testers around the true faces in the image. Since what exactly constitutes a “full frontal” face can be subjective, this can lead to inaccuracies when generating ground truth rectangles, which in turn can result in inaccurate and therefore less meaningful TPR values. In order to avoid this, one can introduce rules when generating the ground truth rectangles in an attempt to reduce subjectivity.

In the pursuit of reducing the aforementioned subjectivity and increasing rigorousness, criteria were applied when creating the ground truth rectangles. If one considers a straight on face to be 0° then anything that is estimated to be within a range of $-45° \leq \theta \leq 45°$ rotation in the $x$ or $y$ directions was accepted as a “full frontal” face.

In addition to having a rule for creating the ground truth, there must also be a rule for whether a detection frame is accepted as having detected a ground truth, or rejected. We decided that this rule must incorporate both the position of the detection frame relative to the ground truth rectangle, and also the relative sizes of the two frames.

**Definition 1.** *Let there be ground truth frame $G$ and detector frame $D$ contained within an image of pixel dimensions $X$, $Y$. $G$ has centre $(x_g, y_g)$ and area $A_g$, whilst $D$ has centre $(x_d,y_d)$ and area $A_d$). $D$ is a True Positive w.r.t $G$ if and only if:*

$$\frac{\sqrt{(x_g - x_d)^2 + (y_g - y_d)^2}}{\sqrt{XY}} \max\left(\frac{A_g}{A_d}, \frac{A_d}{A_g}\right) \leq T \qquad (1)$$

*where $T$ is some threshold determined by the tester.*

Note the $\sqrt{XY}$ term makes the distance factor scale invariant w.r.t the size of the image.

This rule ensures that the distance between the center of the ground truth frame and the detector frame is small. Multiplying this distance by the difference in scale of the two frames encodes our belief that detector frames of similar scales to ground truth frames are more likely to constitute genuine detections.

Our Threshold value of $T = 0.125$ was chosen through experimentation in a way which detected frames we felt constituted a “true” detection. Note that this is also subjective and thus subject to human error. This highlights another practical difficulty in assessing the TPR accurately.


Figure 1: *Here we show the results of face detection run on Dart images 14 (top left), 15 (top right), 5 (bottom right), 4 (bottom centre), and 13 (bottom left). The ground truth rectangles are represented in green, True Positive detections in purple, and False Positive detections in blue.*

The True Positive Rate (TPR) is defined as the ratio of True Positives to the true total number of faces in the image. We can calculate the TPR as:

$$TPR = \frac{TP}{TP + FN} \qquad (2)$$

A TPR of 1 would indicate that every face in the original image is detected by the program. Note that this can be achieved in any detection task by simply setting your detector to consider every possible frame a positive. Clearly then, the TPR alone is not a good measure of the quality of a detector.

The F1 Score is a measure of accuracy of a detector. It can be calculated as:

$$F1 = \frac{2TP}{2TP + FP + FN} \qquad (3)$$

After running our Viola Jones detector on all test images, we found that in total, our detector returned 18 True Positives, 30 False Positives, and 1 False Negative. Thus, using the equations above, the TPR of our detector has been calculated to be $\frac{18}{19} \approx 0.95$. The F1 score of our detector has been calculated to be $\frac{36}{67} \approx 0.54$.

| n | TP | FP | FN | TPR |
|---|---|---|---|---|
| 5 | 11 | 3 | 0 | 1 |
| 15 | 0 | 4 | 0 | 0 |

Table 1: *The number of True Positives, False Positives and False Negatives, with calculated values of the True Positive Rates for dart5 and dart15.*

## 2 V-J Dartboard Detection

The False Positive Rate is defined as the ratio of false positives to the total number of negatives:

$$FPR = \frac{FP}{FP + TN} \tag{4}$$

As shown in Fig.2, the TPR of V-J in training is 1 throughout since it does not miss any dartboards in the images, leading to zero false negatives. Before training, the detector will simply pick out every possible frame position and size in the image and assume they are dartboards. This means that it is impossible for the detector to pick out any true negatives in the images, resulting in a FPR of 1. The FPR subsequently falls over the course of training, as the detector "learns" to reject some images, allowing it to find True Negatives. The final FPR from training is 0.014937.

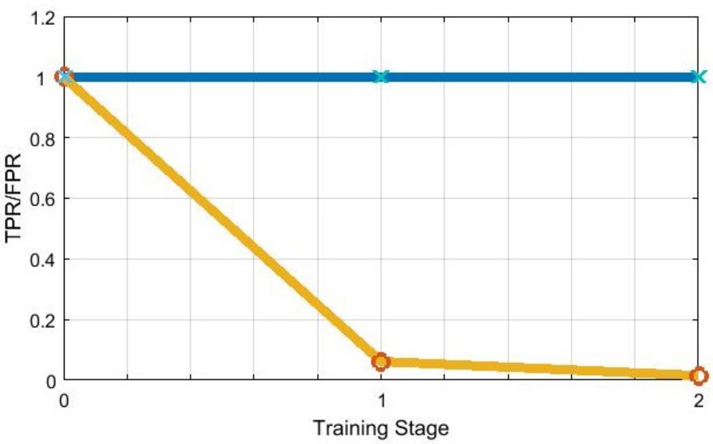


Figure 2: *The progression of the FPR (orange) and TPR (blue) of the V-J dartboard detector in training.*

Initially, we trained a Viola-Jones dartboard detector using the parameters provided on the tasksheet. However, when evaluating this detector, it was noticed that the detector was producing a large amount of False Negatives. It was hypothesized that some of these False Negatives we a result of the dartboards being viewed at an angle too extreme for the detector. In response to this, the V-J detector was retrained, replacing the old maximum angle of rotation in the $x$ and $y$ directions (0.8 radians) with the larger value of 1.2 radians. This had the desired effect of reducing the amount of False Negatives, however the detector was still producing a number of False Negatives which were not at extreme angles. As a wider range of rotations were being considered when generating our positive training set, we also increased the total number of samples in this set from 500 to 1000, this helped incorporate this wider range of information.

The plan had been to run V-J first in our Task 3 detector, however running something with a high FNR first in the process would create a relatively ineffective detector. Hence the V-J detector was retrained with the Maximum False Alarm Rate parameter increased to 0.075, giving a far greater number of False Positives, but consequently fewer false negatives. We considered this acceptable for a first round detector since the other detection methods would be able to eliminate False Positives later in the process.

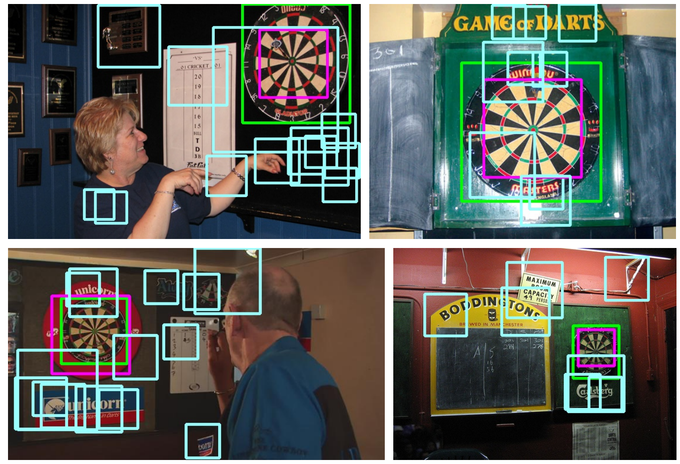

Figure 3: *Here we show Dart images 0 (top left), 1 (top right), 2 (bottom right), 3 (bottom left) when analysed by our Viola-Jones dartboard detector. The ground truth rectangles are represented in green, True Positive detections in purple, and False Positive detections in blue.*

| n | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | T |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| TP | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 6 | 1 | 3 | 1 | 1 | 1 | 2 | 2 | 26 |
| FP | 13 | 9 | 14 | 7 | 9 | 10 | 19 | 19 | 22 | 23 | 46 | 6 | 5 | 16 | 40 | 8 | 266 |
| FN | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |

Figure 4: *This table shows the TP, FP, and FN scores of the V-J dartboard detector when run on the test images. These scores will be used to calculate the F1 score below.*

By equations (2) and (3), the F1 score of our Viola-Jones dartboard detector is F1 = 0.16 and the TPR = 0.95.

- The detector had TPR of 1 in training, but in testing it came out to be 0.95. This is due to different lighting conditions and occlusion which occur in the test images, but not in the positive training set, where only viewing angle was taken into account. We conclude that the TPR and FPR values during the training stage are not particularly useful estimates of how well the detector will perform on images outside of the training set.
- The low F1 score of 0.16 in testing is a direct result of the alteration of the Maximum False Alarm Rate. It would be considered a relatively poor performance were we to be running this detector as it is, however as previously mentioned, it is acceptable for our purposes.

# 3 Integrating Shape Detection

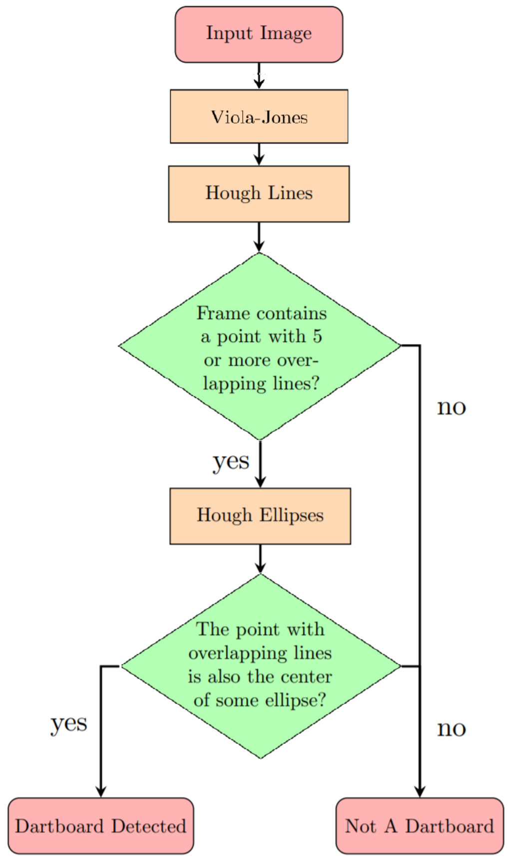


Figure 5: *Flow chart outlining the decision making process of our dartboard detector.*

- Our Viola-Jones detector is purpose built for running first; it is the fastest of the our detection methods and has an very low FNR. It outputs potential frames.
- A Hough line transform is applied to the thresholded gradient image of each frame. Every point on the Hough space above a certain threshold value is considered to be a true line in the image. If our frame contains a point with 5 or more overlapping lines, then the point with the most overlapping lines is considered to be a potential dartboard center. This algorithm was computationally quick, and proved incredibly effective for identifying false positives. This algorithm was able to identify the majority of the false positives. Note in Fig.6 the Hough spaces are quite sparse, this is because data points were plotted at specific angles for each point in the thresholded gradient image, according to their values in the gradient direction image.
- Lastly, these points are run through Hough Ellipses as potential centres. This provides an extra check in determining whether we have a dartboard centre, and also gives information about the size of that dartboard. It is run last in the interests of performance, since it is the most computationally demanding method.

To demonstrate the merits and limitations of our detector we have chosen dart8 and dart10 respectively. In dart8 the detector has located (with no FP's) two dartboards; both different scales, orientations, lightings, and with one of the dartboards partially occluded. Also, the detector provided reasonably good estimates of the size of each dartboard, though the partial occlusion of one board has led to a slightly inaccurate estimate.

Meanwhile, dart10 managed to perform in all of these circumstances bar occlusion for two dartboards, but has produced a FN for the third dartboard in the image. Also, the estimated radius for one of the dartboards has been scuppered by a larger ellipse around the dartboard, which due to its size will have likely have had more "votes" in the ellipse Hough space, resulting in the ellipse detector taking this radius as the size of the dartboard. This phenomenon can be seen in multiple images, and is not something we had anticipated.

| n | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | T |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| TP | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 2 | 1 | 1 | 1 | 2 | 1 | 19 |
| FP | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| FN | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |

Figure 6: *This table shows the performance of the cascade dartboard detector on the sample images.*

By definitions (2) & (3), over all sample images our dartboard detector has: TPR = 0.95 and F1 = 0.97. Notably, the increased number of FP's produced in retraining the V-J detector have not percolated through the system. This shows that retraining the V-J detector with a high false alarm rate has had the desired effect; keeping the number of FN's low whilst not increasing the FPR of the overall detector.

Figure 7: *dart10 (top left) & dart8 (top right) when analysed by our dartboard detector. There are ground truth rectangles (green) and True Positive detections (purple). Below are images from four stages of the detection process for selected frames, from left to right: The thresholded gradient image, the Hough line space, detected lines on the frames, and the final detections.*

# 4 Improving the Detector

Our Task 3 (T3) detector had two obvious weaknesses; the FN produced in dart10, and a lack of robustness in size estimation. We focused on reducing the FN count of our detector on the grounds that a detection failure is a greater issue than imperfectly predicting a detected dartboard's size.

To reduce the possibility of FN's, another potential frame detector was required to run alongside the V-J detector, hopefully picking up any dartboards the V-J missed.

After consideration, it was decided that the new detection method should utilise the distinctive colour palette of dartboards. To do this, a means of recognising specific colours in the image was required.

When performing colour recognition, it is advantageous to operate in the HSV colour channels, as opposed to RGB. The prevailing logic to this is that it separates color hue information from intensity, making it possible to easily look for a certain range of hues and saturations whilst not having to concern ourselves with the lighting in the image, which should only affect the Value channel. This results in more robust color thresholding over simpler parameters. Through experimentation, we found ranges in HSV to pick out dartboard colours, though these are by no means definitive.

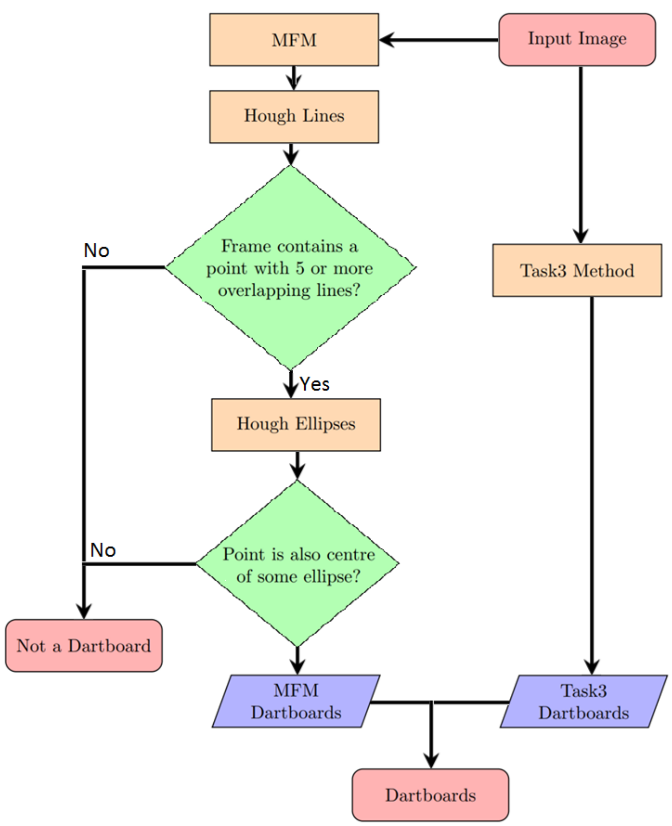


Figure 8: *Updated detection algorithm with MFM integrated.*

The process of our new dartboard detection algorithm with the new "McAdam-Freud McCarthy" (MFM) frame detection method is as follows:

1. Run the T3 detector and collect detected dartboards.
2. Separately, pass a n×n kernel (n odd) over the image.
3. At each position, test if the kernel contains at least one pixel of green, red, and cream. If this is true, then a pixel is turned to white on a blank image at the coordinates of the centre of the kernel (black image is the same dimensions as the test image).
4. Cluster analysis is performed on this image by averaging the positions of coordinates within a certain distance of each other, and a pixel is plotted at the average position of each cluster.
5. Frames of multiple sizes are placed around each centre.
6. Run a Hough line detector followed by a Hough Ellipse detector on the frames (similarly to T3 algorithm). Keep any frames which pass both detections.
7. The frames are then compared with those of the T3 detector. If any new frames overlap with T3 detector frames, the new frames are assumed to be detections of the same dartboard and discarded.

| n | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | T |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| TP | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 3 | 1 | 1 | 1 | 2 | 1 | 20 |
| FP | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| FN | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Figure 9: *This table shows the improved performance of the dartboard detector with the MFM detection method when run on the sample images.*

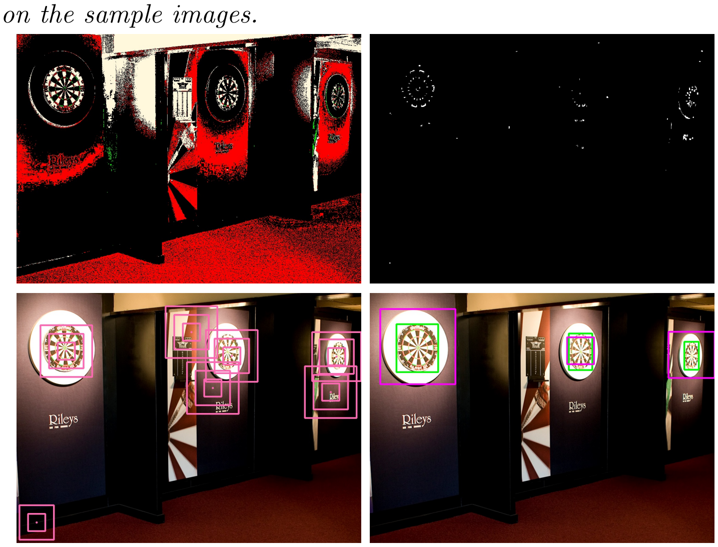

Figure 10: *Visualization of dart10 pixels considered red, green and cream (top left), points plotted by kernel sweep (top right), cluster centres and generated MFM frames (bottom left), and final detected dartboards (bottom right).*

Fig.10 shows that our MFM detection method has eliminated the FN, as planned. This gives the final detector TPR = 1 and F1 = 1. This new method does have the disadvantage of not being able to work on greyscale images, however the majority of images are taken in full colour these days so this should not prove to be a major problem. Comparing this detector to T3, we see something of a trade off between speed and accuracy, since the new version is slower, but produces 0 False Negatives.