Department Of Computer Science
University Of Bristol

The Dartboard Challenge

MSc in Robotics

Albert Jose Potams and Sean Wheeler

December 4, 2017

1 Subtask 1

1.1 Part A

(a) Dart4.jpg (b) Dart5.jpg (c) Dart13.jpg

(d) Dart14.jpg (e) Dart15.jpg

Figure 1: Detections from the frontal face cascade, red boxes indicate a true positive detection and turquoise boxes indicate a false positive detection.

1.2 Part B

For dart5.jpg the TPR $= \frac{11}{11} = 1$. For dart15.jpg TPR $= \frac{2}{3} = 0.3\dot{3}$.

When the cascade runs, the different classifiers search for certain features and if they do detect the features they return a positive value. In some cases it may be that a classifier returns a positive value and detects a certain feature when in actuality the particular feature it was trained on is absent, in some cases this may even happen when the sub-window is indeed focused on a face for example. For this situation it is difficult to say whether this has occurred and is a true positive detection or not. In the case of dart15.jpg, it is difficult to assess whether there are really any ground truth samples to detect, this is as the cascade is built to classify frontal faces. In this image all of the faces are side on and so may not truly be considered ground truth. However despite this, the cascade detects one face fairly accurately in figure 1e) and one additional detection is of the upper half of a face. Whether this can be considered a true positive or not is debatable. Arguably it is not detecting what it is trained to, but one could equally say it is simply detecting the same object from a different view-point and that this demonstrates a robust cascade of classifiers.

The TPR is defined in equation 1. It does not take into account false positives, this means that if the threshold for detection is low and the classifiers return positive values very easily then every single sub-window could technically be classified as a positive detection. If this were so it would ensure a TPR of 100% but also that the classifiers are essentially useless.

The F1 score is defined in equation 2. In order to make it meaningful and accurate, the underlying variables must be too. Essentially, the determination of all of the underlying variables depends on whether a detection is made and whether it is sufficiently similar to the ground truth to be classified as a true detection or not. For this program tolerance values have been implemented, derived from both the width and the height of the detection frame. If the top left corner and bottom right corners of the detection frame are both within the tolerance range of the same corners of the ground truth frame then the detection is classified as a true positive detection. If not then the detection is classified as false positive. If any of the ground truth frames do not have a corresponding detection frame then this implies a false negative.

2 Subtask 2

2.1 Part A

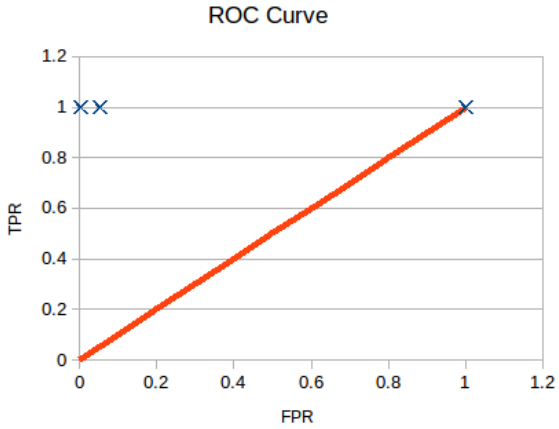
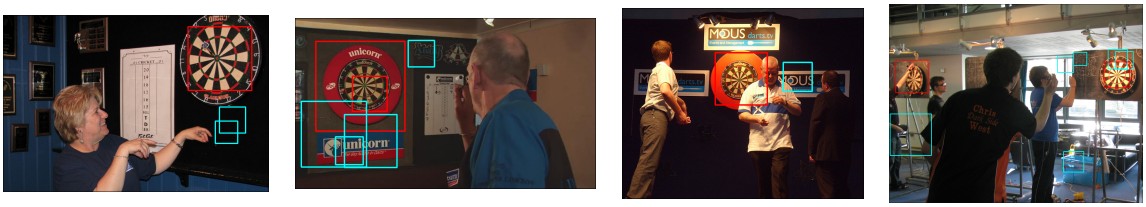


Figure 2: TPR vs. FPR for the three stages of the dartboard classifier.

Figure 2 shows that for all stages of the classifier, the reported TPR is 1. For the first stage, the FPR is also 1, this means the stage reports a random classification which is essentially useless. For the second and third stages, the FPR decreases to lower values, indicating that successive stages of the classifier have improved in their ability to detect true negatives.

2.2 Part B



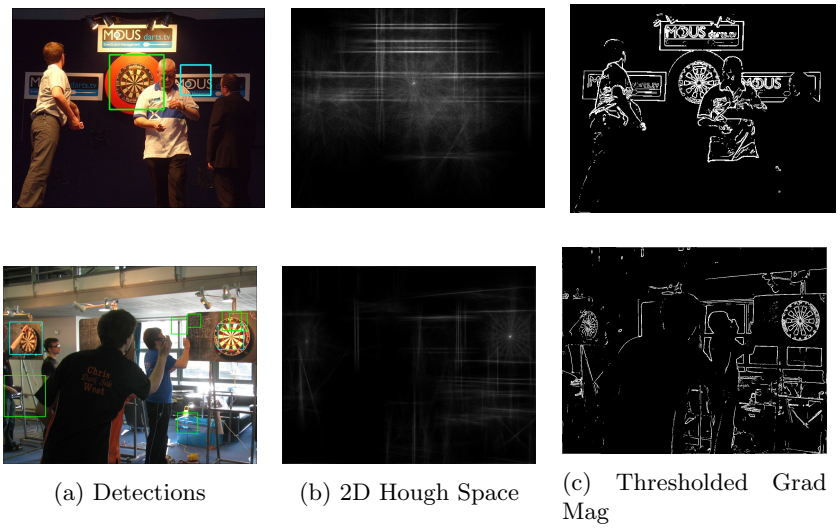
(a) Dart0.jpg (b) Dart2.jpg (c) Dart6.jpg (d) Dart8.jpg

Figure 3: Detections from the dartboard cascade, red boxes indicate a true positive detection and turquoise boxes indicate a false positive detection.

- Using equation 1 the mean **TPR** across all 16 images was calculated to be **0.92**. Likewise, using equation 2 the mean **F1 score** was **0.56**. The **precision** (equation 3), was **0.45**.
- Compared to fig. 2 the true positive rate is close to what was reported at all stages of the classifier, with most dartboards being detected. Due to the number of true negatives not being readily available, we are unable to compare final results with the ones generated by the cascade training program. However it is reasonable to assume since all areas of the image are being sampled at different scales, and there are relatively few positives at all compared to the total area, that the number of true negatives is fairly high and the FPR would be on the order of that reported by the second and third stages of the classifier.
- Without comparison to classifiers built in the same way for different objects, it is hard to quantitatively say whether the TPR and FPR generated by cascade training are good measures of how the system will perform on real testing data. However, in this case we can say that the reported rates are reasonably similar to what the training data shows, perhaps with exception to the TPR. Although the absolute difference is small (0.08) this difference implies that not all dartboards will be detected, whereas the original measures in fig. 2 suggest that all dartboards will be detected without fail.

3 Subtask 3

3.1 Part A



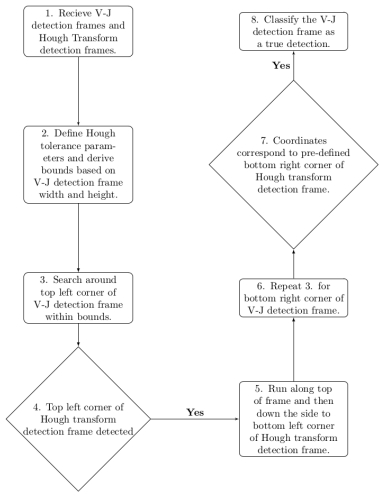
(a) Detections (b) 2D Hough Space (c) Thresholded Grad Mag

Figure 4: V-J (green) and V-J+Hough Transform (turquoise).

3.2 Part B

Combining the Viola-Jones method with a Hough circle transform produced a mean F1 score of **0.85** across the set of images. This is quite clearly a significant improvement over the previous detector (F1=0.56), however the disadvantage of the process we have used to combine the two methods means that the true positive rate has decreased from 0.92 to **0.88**. This is because our method will take all V-J detections and then simply discard ones which do not have a sufficiently matching circle in the Hough space, meaning if only one method fails to detect a dartboard then the detection will be classified as a negative. Overall however, due to the increase in the F1 score we can say this is an improved detector than one simply making use of V-J detection. The precision of the method has shown a large improvement from 0.45 to **0.81**, this is due to large numbers of false positives being discarded by the method, as opposed to relatively few true positives.

3.3 Part C



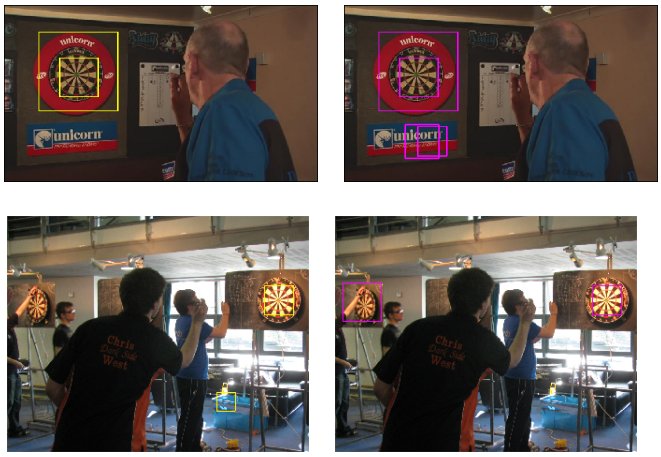
- The combination of evidence is based on there being a certain similarity in frames, both in position and scale. Due to the consideration of opposite corners.
- This similarity is based on a tolerance value which was chosen under consideration of the evidence being generated by both methods.

4 Subtask 4

4.1 Part A

- We have used the additional Hough Space approach, using the intersections of lines rather than the centre of circles as our additional method.
- Our rationale is based on the fact that most of the dartboards share the uniform line characteristic, where the intersection is at the centre of the dartboard.
- Extracting this should help us reduce the number of false positives present in our detections. As if this information is not present we can reasonably assume it is not a dartboard.
- We will compare the centre of the circles detected in the previous method with the point of the line intersection, if the points are approximately close, we will assume this is a true positive.

4.2 Part B



(a) V-J, Hough Circles and Line intersections

(b) V-J and Hough Circles only

Figure 6: Dart2.jpg (top) is a good example. Dart8.jpg (bottom) is a bad example.

4.3 Part C

Overall the values for the F1 score decreased from 0.85 to **0.83**. This is because although the method was successful in reducing the total number of false positives, it was only marginal, with 1 false positive being eliminated. And a further two true positives were missed by the method causing an overall decrease. The TPR decreased from 0.88 to **0.85**. Again this was caused by a reduction in the number of true positives (and hence increase in FNs). The precision decreased as well overall to **0.79**, which again indicative of the overall drawbacks of the method, that in this case it eliminated more true positives than false negatives. This approach could be further improved to be accurate if the tolerance values for the maximum distance between the intersection centre and circle centre were further refined. The circle centres for the method were picked in the final function for detection before certain centres were removed if they were too close to another centre. A simple error that increased a few false positives.
Finally, it was observed that true positives located far away in the image or not clearly with no clear line intersections were not picked up.

A Appendix

$$TPR = \frac{TP}{TP + FN} \tag{1}$$

$$F1 = \frac{2TP}{2TP + FP + FN} \tag{2}$$

$$Precision, PPV = \frac{TP}{(TP + FP)} \tag{3}$$