Introduction to Vision and Robotics
Vision Practical: Coin Counter

Dylan Angus, Matthew Martin

October 26, 2016

1 Introduction

The purpose of this practical is to develop a program in Matlab that recognises and classifies several objects in an image. These objects can be coins or other small items, and the program must segment the image, identify each of the objects, and output the total value (in pounds and pence) of the objects in the image.

All of the images are taken from a downward facing camera viewing a scene containing the objects on a static background. We were provided with a set of 14 sample images on which to train our classifier (see Figure **??** for an example).

The following are the objects and associated values that may or may not be present in any given image:

- one and two pound pieces
- 50, 20, and 5 pence pieces
- washer with small hole (75p)
- washer with large hole (25p)
- angle bracket (2p)
- AAA battery (no value)
- nut (no value)

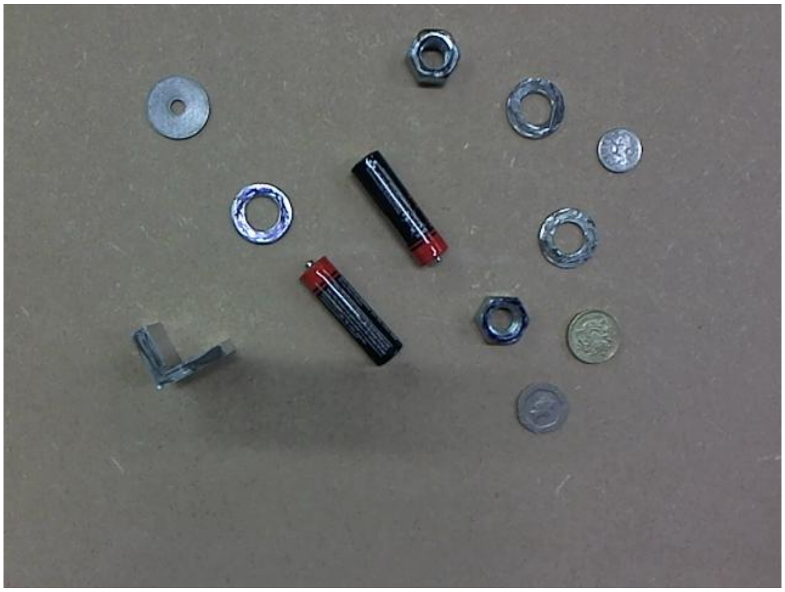


Figure 1: This is one of the test images given to train the classifier.

We approached this problem by dividing it into three distinct stages: background segmentation, object detection, and object classification.

2 Methods

2.1 Background Segmentation

Creating a reliable algorithm that would clearly segment the background from the objects of interest in the image was the most time-consuming and challenging section of this project. We tried several different methods of background segmentation, to varying degrees of success. We ended up choosing median filter thresholding as our most successful method.

2.1.1 Naive thresholding

Our algorithm for creating a naive threshold can be described by the following steps:

1. Attain the median values for each of the three color channels, r, g, b in the given image

2. For each pixel, if that pixel's color values are ±20 from the median, label it as a background pixel. Else, label it as an object of interest.

This method has a few advantages. It is fast, as it only requires two passes over the entire image, and there are no computationally expensive operations inside of the loop. It is simple and easy to understand. However, this method fails to accommodate for shadows in the background. It also needs to be tuned specifically to the image (the range of ±20 from the medians was chosen by trial and error). Even after careful tuning, this algorithm still miss-classifies some pixels. See Figure 2a for an example of the output of this method.

2.1.2 Adaptive thresholding

Adaptive thresholding, as opposed to naive thresholding, generates a unique threshold value for a set of sub-images inside the given image. This is meant to allow for shadows to fall on the background and still be classified as background since, the threshold is a more localized value.

This was a fairly successful method, but still had its share of disadvantages. Adaptive thresholding highlighted the edges around some of the objects rather than the objects themselves, but for others, identified the body of the object correctly. This inconsistency would cause problems when trying to classify the object. However, we never had any problems with background shadows when using this method. See the results in Figure 2b.

2.1.3 RGB normalization

This algorithm is meant to reduce background shadows as well by normalizing the intensity of a pixel color. It adjusts the r, g, b values based on the following division:

$$r = \frac{r}{r+g+b}, \quad g = \frac{g}{r+g+b}, \quad b = \frac{b}{r+g+b}.$$

Then, a tight threshold can be created based on the now very similar background values.

RGB normalization has a lot of advantages. Like adaptive thresholding, it is effective at ignoring shadows in the background. It also runs very quickly since the normalization takes advantage of Matlab's efficiency in vectorized operations, and then requires only one pass over the pixels to threshold. It produces fairly consistently good results, rarely classifying background pixels as an object. However, it often misses a couple objects in each image, which is unpredictable and would be problematic for trying to get an accurate money total for all objects in the scene. See Figure 2c for an image segmented by RGB normalization.

2.1.4 K-means clustering

We also tried using a K-means clustering algorithm to segment the background from the foreground. This algorithm is meant to cluster all of the background pixels into the same class and then cluster all of the objects into the same class.

We saw limited success using this strategy, as it often fails to classify shadows in the background as part of the background. It is also fairly slow, as it has to run for several iterations before the clusters converge. See the results from this method in Figure 2d.

2.1.5 Mean-shift segmentation

Mean-shift segmentation is a commonly recommended method for background subtraction, because, like adaptive thresholding, it localizes its segmentation to subsets of the pixels in the image. The algorithm searches for local maxima in the data and identifies that as an approximate for the background.

The performance of this algorithm varies tremendously throughout different testing images. In some images, it separates out the background almost perfectly, while in other images (see Figure 2e) it localizes the maximum as a background shadow, and classifies the shadow as the background. It is also a time consuming algorithm. Ultimately, we could not reduce the inconsistency in performance.

2.1.6 Median filter thresholding

Median filter thresholding involves using the entire dataset of images for background segmentation. This algorithm calculates the median r, g, b values for each pixel throughout all of the images in the set. Then we are able to compute a pixel-by-pixel threshold for the image being segmented. We knew that the median r, g, b values would correspond to the background color for that pixel location, so we put a threshold on the sum of the absolute values of the differences between the median r, g, b value at a pixel and the actual

r, g, b value at that pixel. Here is the threshold represented mathematically, for a single pixel and a threshold T:

$$abs(r_{median} - r) + abs(g_{median} - g) + abs(b_{median} - b) < T$$

This algorithm consistently performed very well. It ran fairly quickly, especially after taking advantage of Matlab's vector operation optimizations. It always classified background shadows as part of the background, and usually found most of each object. We decided to use this method for our background segmentation algorithm. See Figure 2f for a sample output image.

2.2 Object Detection

We tried a couple different methods for detecting each object from the segmented image. The success of these methods is largely dependent on how well the background is separated from the original image. We settled on detecting boundaries in the image as the most reliable method to recognise discrete objects in the segmented image.

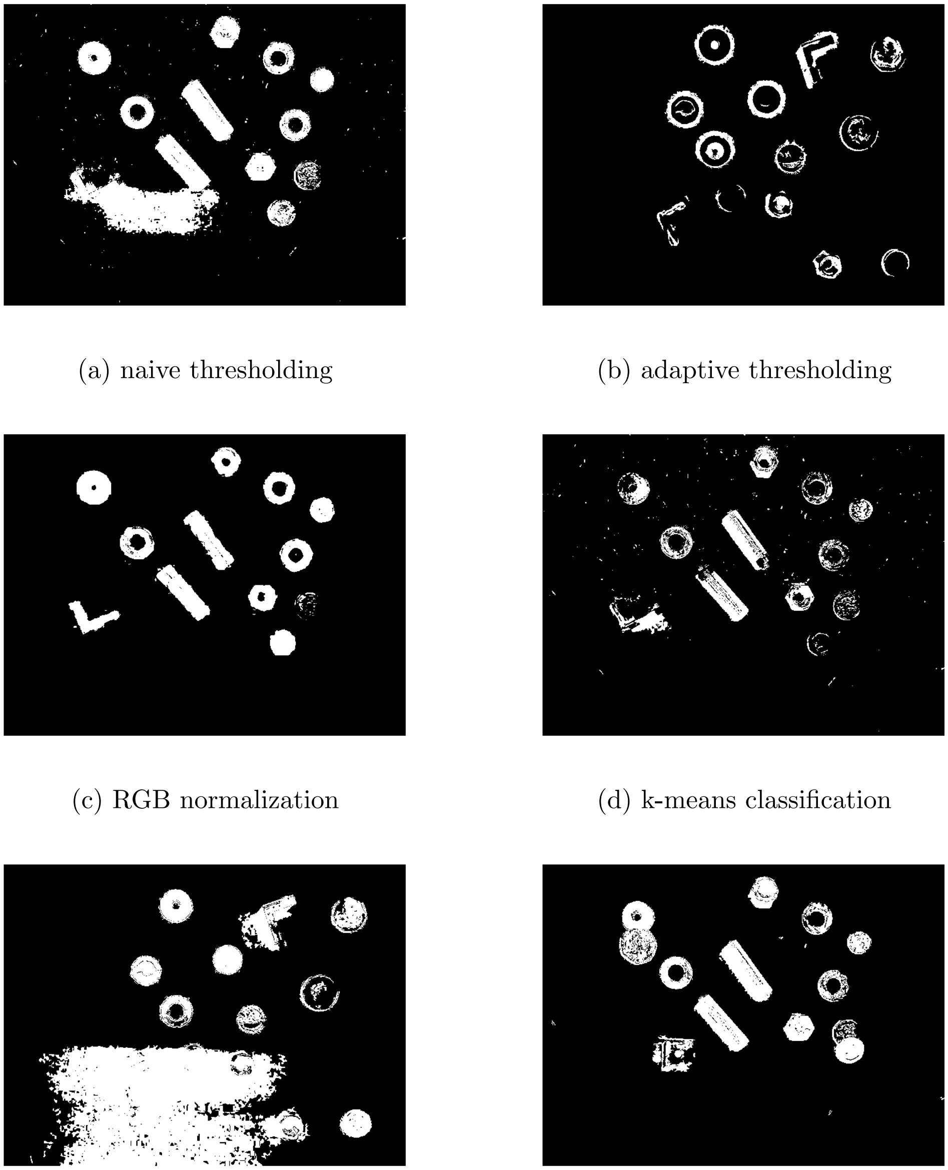
2.2.1 K-means clustering

We attempted to use k-means clustering in order to separate out the different objects in the image. The idea was to have each cluster correspond to an object in the image. However, we realized that this would not be practical to use generally because the k-means algorithm requires the number of clusters as a parameter. Since we do not know how many objects are in each image before processing it, we do not have this information for the classifier.

2.2.2 Boundary detection

Boundary detection works extremely well to separate out each object. The only issue is that sometimes the background segmentation breaks up an object into a few pieces. We solved this problem by doing some pre-processing on the image before passing it to the boundary detector. The pre-processing algorithm is structured as follows:

- For each pixel in the image:
- If the pixel is white (part of an object), and the pixel that is 10 rows below it is also white, then connect the two pixels



(a) naive thresholding

(b) adaptive thresholding

(c) RGB normalization

(d) k-means classification

(e) mean-shift segmentation

(f) median filter thresholding

Figure 2: Here are the output images for the six methods of background segmentation that we tried.

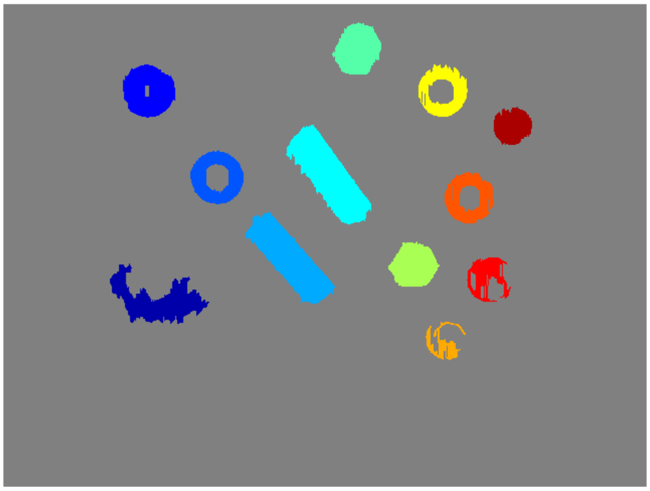


Figure 3: Here is an image after background subtraction and boundary detection. Each color corresponds to a unique object as identified by its boundary.

This algorithm effectively joins up objects that are meant to be together, thus making it so they share a boundary. This image is then passed into a boundary detector, and then we return a logical matrix that is 3-dimensional: rows, columns, and a layer for each object in the image. See Figure 3 for the output of this process.

2.3 Classification

The classification process consisted of two main parts: feature selection and classifier training.

2.3.1 Feature selection

Feature selection is an extremely important stage in claification. Choosing features that allow the classifier to make clear distinctions between different objects in the image is essential to the success of the classifier. We chose to extract the following features from each object in the image for our classifier:

- Mean r, g, b values

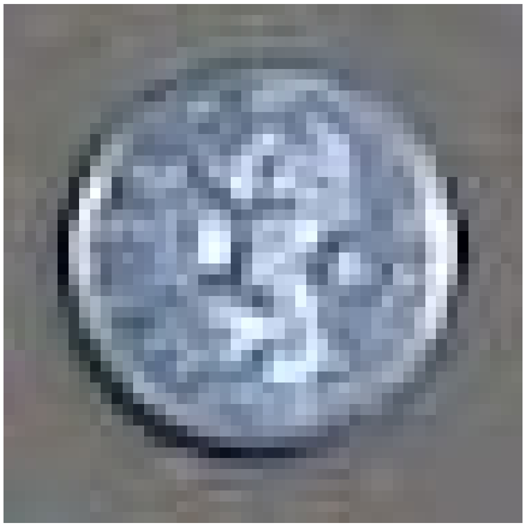


Figure 4: A five pence piece extracted from a training image.

- 1 complex moment (ci1)
- Compactness

These features were calculated given a subimage of the overall image that contains only the object of interest. See Figure 4 for an example of a 5 pence piece that would be analyzed for these features.

The process of deciding to use this feature set relied on trial and error. We attempted several other combinations as well, but achieved worse results as compared to the accuracy of these features. We tried to use simply mean r, g, b values and 6 complex moments. In this instance, the extra several complex moments did not seem to improve performance, so we eliminated them. We tried adding in major axis length and minor axis length, however, this was only useful in separating out the batteries and angle brackets, which were already being classified fairly accurately. We then tried integrating SURF and FAST features into the analysis, but the low resolution of the segmented objects caused these features to have a very little influence on the overall classification. Further, the acquisition of these two features took longer than any others, so the extra computation time did not justify the minimal/lack of a benefit. It turned out that keeping the features as concise as possible yielded the best results.

2.3.2 Classifier training

Upon testing the multivariate gaussian classifier with all the features in the feature vector grouped together, we realized that there was not enough training data to properly approximate the covariance matrix. Thus, we then chose to approximate the distribution by using a Naive Bayes classifier instead. By assuming independence between the group of RGB means and group containing the compactness and c1 moment together, we used a multivariate gaussian distribution to approximate each of their conditional distributions and then found the class which maximised the product of their conditional probabilities and prior probabilities of the classes. Doing this the performance rose drastically.

3 Results

In order to assess the accuracy of our classification model, we split the data into a training and a testing set, giving 75% to training and 25% to testing. We calculated *precision*, *recall*, and F_1 values for each class. These two measures are widely used to assess the performance of classification models and are calculated as follows (where $tp = TruePositives, fp = FalsePositives, fn = FalseNegatives$):

$$precision = \frac{tp}{tp + fp}, \quad recall = \frac{tp}{tp + fn}, \quad F_1 = \frac{precision * recall}{precision + recall}$$

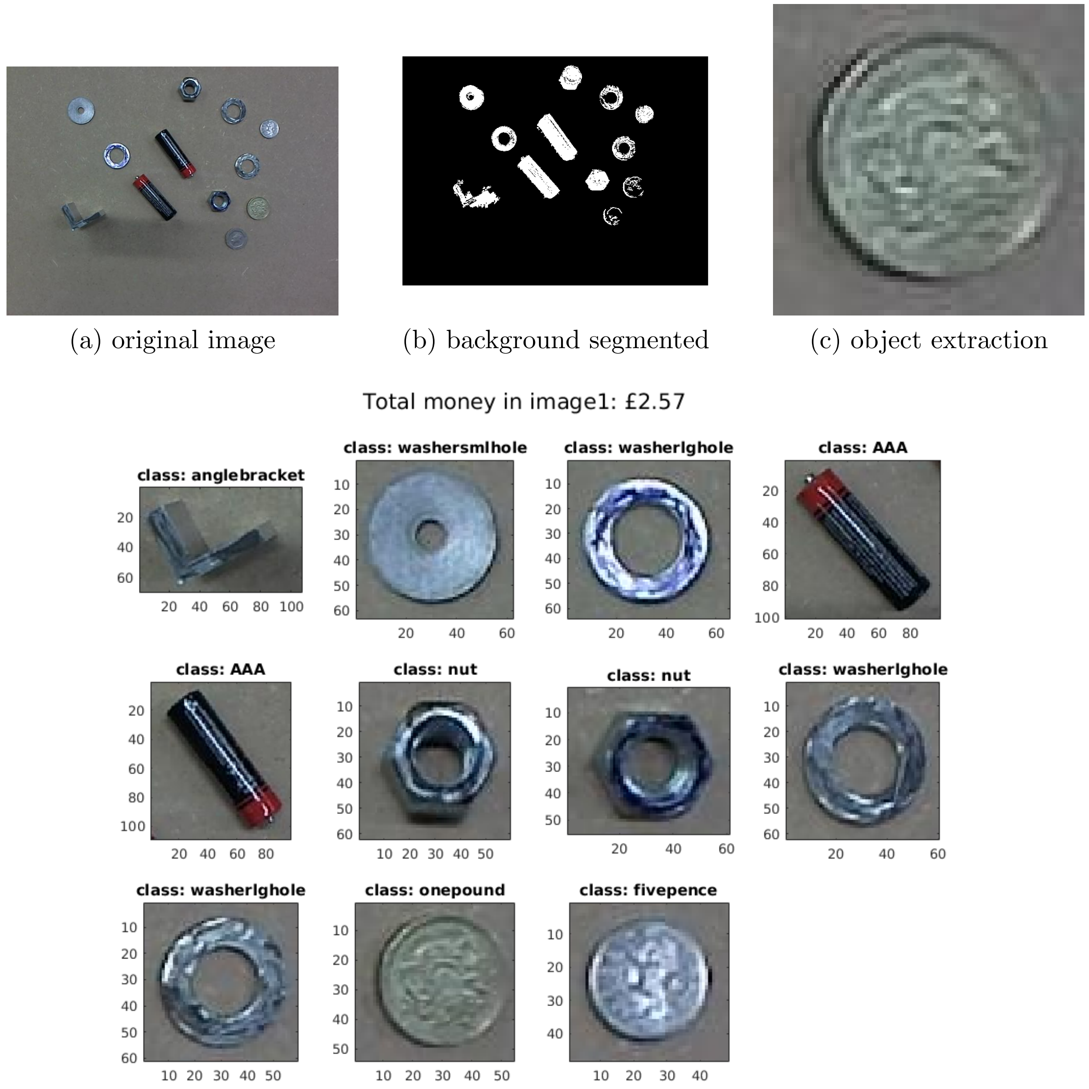
We calculated average precision and recall across each of the ten classes in the testing data and achieved the following results:

$$precision = 0.739,$$
$$recall = 0.731,$$
$$F_1 = 0.715.$$

We generated a confusion matrix from our testing data. This can be seen in Table 1. Figure 5 shows the output of our program at each stage of the process, as described in the Methods section.

Table 1: Confusion matrix for data from testing classification. Zeros were omitted for readability.

	AAA	aBracket	50p	5p	Nut	1Pound	20p	2Pound	WashLgHole	WashSmHole	NotSure
AAA	2										
aBracket		2									
50p			1				1				
5p			1	3							
Nut					2						
1Pound						2					
20p		1					2		1		1
2Pound							1	2		1	
WashLgHole							1		8		
WashSmHole					2					3	
NotSure											



(a) original image

(b) background segmented

(c) object extraction

(d) classification

Figure 5: This shows the overall pipeline of our program. The original image is segmented, objects are extracted, and then classified.

4 Discussion

Overall, we are satisfied with how our classifier performs. That said, even though it has many strengths, it is also hindered by a few limitations.

4.1 Strengths

4.1.1 Fast

Our model performs classification very quickly, especially after the removal of SURF and FAST features. Training the model takes 2.47 seconds. Classifying images takes 1.58 seconds for 9 images, which comes to about 0.18 seconds per image classified.

4.1.2 Accurate

Our high *precision*, *recall*, and F_1 measures indicate the strong performance of our classifier.

4.1.3 Conservative

By adding in the $NotSure$ class, we allowed the algorithm to decline to objects that were unclear. We believe this is a strength because it lowers the number of mis-classifications, which is better in many real world applications than trying to classify everything and having more mis-classifications.

4.2 Limitations

4.2.1 Training data

One issue with our model is the amount of data used to train it. This could have been improved if we were able to capture more images of the scene, but we did not have time to do this. A larger sample size for training would undoubtedly result in better accuracy measures.

4.2.2 Background restrictions

Our method of background subtraction relies on a dataset where all the images have the same background. If there were to be varied background in the data, then the median filter would only filter out the background pixels that happen to be at the median of the different background colors, and miss all of the others. This is not a problem for this particular project since we know all of the images have the same background, but in more general applications, the median filter would not work.

Appendix

```
code can go here
```