SPS Coursework 2 Report

Daniel Keitley (dk14988)

1 Introduction

In this report I attempt to describe, analyse and explain my results obtained from classifying images to recognise T, V and S handwritten characters.

2 Feature Selection

To obtain features of the 3 characters, I used a step-wise feature selection approach, using the Fourier transforms of the 3 characters. In the discrete form, the Fourier transform is given by:

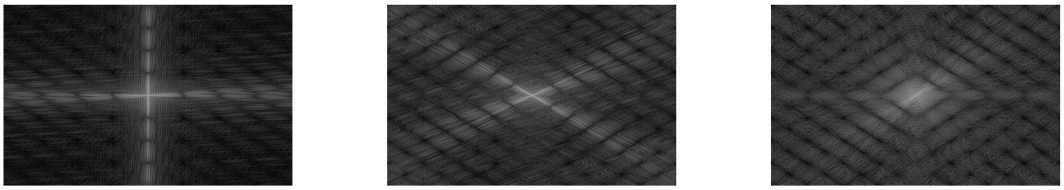
$$F(u,v) = \frac{1}{N^2} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x,y) \left[\cos\left(\frac{2\pi(ux+vy)}{N}\right) - j\sin\left(\frac{2\pi(ux+vy)}{N}\right)\right]$$

where the $f(x,y)$ term represents the intensity of the image at the position (x,y). The value of $F(u,v)$ is given by a sum of sines and cosines with spatial frequencies ranging from 0 to the nyquist frequency, weighted by the intensity of the image at points corresponding to these spatial frequencies. The spatial frequency refers to the frequency across the image for which the brightness modulates[1]. Therefore, high magnitudes within the high frequency ranges of the Fourier domain corresponds to edges within the image, as this relates to a rapid change in image intensity. For the letter T, the Fourier space consists of concentrated horizontal and vertical lines of high magnitude (see figure 1a). The horizontal line for example shows high magnitudes for very low frequencies in the v (vertical) direction and for very high frequencies in the u (horizontal) direction, thus symbolising a vertical edge. For the letter V, the Fourier space consists of concentrated regions in the left and right diagonals (see figure 1b). This, by a similar reasoning, represents the diagonal edges in the V character. Finally, the letter S consists of lower magnitudes in all directions as a result of the curvature in the character. There is also significantly high magnitudes in the right-diagonal of the Fourier space due to the left-diagonal component of the letter (see figure 1c).

This reasoning suggests that the power in these regions, given by $\sum_u \sum_v |F(u,v)|^2$, could act as good classification features. In particular, I selected a vertical rectangular region to characterise the horizontal line of the letter T (feature 1, see figure 2a) and a sector in the left diagonal region of the Fourier space to characterise the right-diagonal line of the letter V (feature 2, see figure 2b). I noticed that the diagonal lines of the letter V are closer to vertical lines than to horizontal lines and so the vertical region in the Fourier space (corresponding to horizontal lines) has a lower magnitude for the letter V, than for the letter T, which is very concentrated in this region. Hence this region defines a feature which strongly classifies the letter T. The lower half of the Fourier space was ignored due to the fact image intensity is a real-valued function meaning by the conjugate symmetry property of the Fourier transform, the negative frequency values of the Fourier space is equivalent to the positive frequencies. For my second feature, I captured the right diagonal line characteristic of the letter V. I avoided using the left-diagonal due to the left-diagonal section in the letter S, which gives high magnitudes in the same region. Hence, this

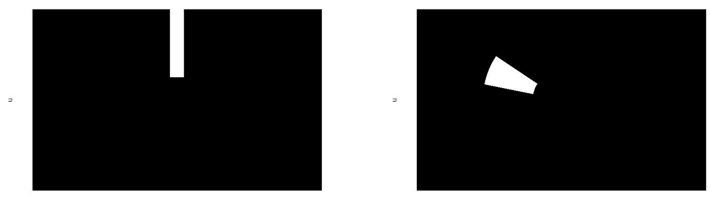
[1](2016). An Intuitive Explanation of Fourier Theory. [online] Available at: http://cns-alumni.bu.edu/ slehar/fourier/fourier.html [Accessed 29 Apr. 2016].

second feature strongly identifies the letter V. Using a sector accounts for variations in the angle of the diagonal lines, and for rotated characters (since rotation of an image results in rotation of its Fourier space). After experimenting with using more features, with the width of the rectangular feature and the sector radius and angle, I saw no improvement in the ability of my features to distinguish between the characters and so I stuck with these two features.



(a) (b) (c)

Figure 1: Typical Fourier transforms for letters a) T b) V and c) S. The horizontal axis represents the horizontal spatial frequency u and the vertical axis represents the vertical spatial frequency v. The frequencies (0,0) is located at the centre of the image.



(a) Feature 1 mask (b) Feature 2 mask

Figure 2: Regions of the Fourier space which contributed to the power calculated for features 1 and 2.

3 Analysis of Feature Selection

- **Informativeness**

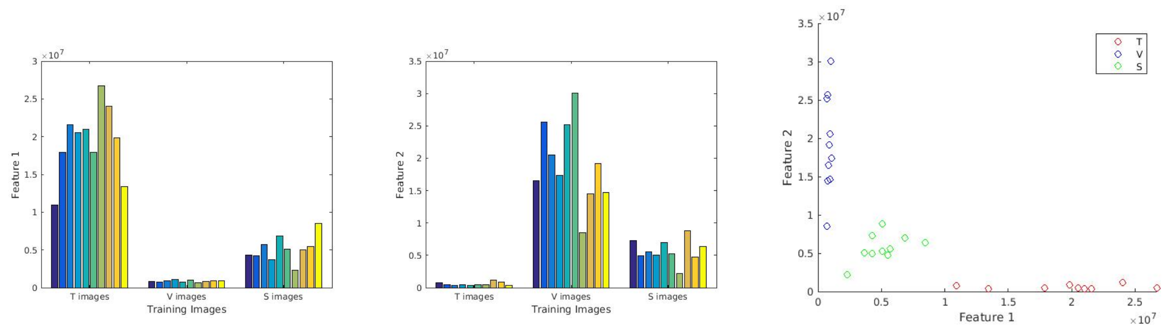
  The graphs in figure 3 confirm the theoretical reasoning above. The feature values of the training images of T, S and V are clearly separated (see figure 3c). The images of the letter T, give high values for feature 1 and very low values for feature 2 (see figures 3a and 3b), corresponding to high magnitudes in the vertical direction of the Fourier space and very low magnitudes in the diagonal directions (as predicted). The images of the letter V conversely give very high values for feature 2, corresponding to concentrated diagonal regions in the Fourier space and very low values for feature 1 (see figures 3a and 3b). The images of the letter S give moderately low values for both features 1 and 2 due to the curvature of the letter which contributes to low magnitudes in all directions of the Fourier space. This suggests that a classifier could be trained to find a clear decision boundary between the 3 letters.

- **Independence**

  It should be noted that the features are not entirely independent. High magnitudes in the vertical rectangular region of the Fourier space, characteristic of a letter T, implies low magnitudes in the diagonal regions, since the letter T has no diagonal edges. Hence there is a negative correlation. After normalising the training data, there is a -0.72 covariance between the feature values. This implies that it is possible, to some degree of accuracy, construct one feature from the other and reduce the number of features through principle component analysis. To construct a feature independent to feature 1 from the Fourier space, it would require taking a region for which S has the highest or lowest power. Otherwise, if T has a very high or low value for feature 2, it also has a high value for feature 1, and so means the features are correlated (either negatively or positively). The same argument can be applied to V. In practice, I found it difficult to construct a feature in the Fourier space with highest or lowest feature values for the letter S. Therefore, it may be possible to construct a more effective classifier by using features outside of the Fourier space. This would give the classifier a different type of information and consequently avoid using redundant features.

- **Simplicity**

  The relationships between features 1 and 2 are very simple and thus easy for a classifier learn. This means that less training data is required to train the classifier. This is particularly useful in this coursework as we are only given 30 training images.



(a) Feature 1 values for each training image

(b) Feature 2 values for each training image

(c) A scatter plot of feature 1 against feature 2 for each training image, coloured according to each letter

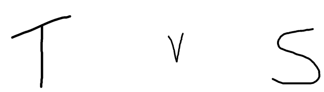
Figure 3

4 Nearest Neighbour Classification

To classify my test images, I then used a nearest neighbour classifier. The feature values for each test image is generated and compared to the feature values of the training images. The test instance is then classified according to the majority class of the K closest feature values, according to the Euclidean distance. I varied this parameter K, using the test data, and found that the classification accuracy peaked at 90% for K=1,2 before converging to approximately 70% (for K$\geq$3). This, however may not be reflective of future data, and if possible, should be tuned according to a cross-validation set, separate to the training and test data sets. For my classifier I chose K=2 in an attempt to avoid incorrect classification of points neighbouring possible outliers in my training data and also to smooth the decision boundaries. The nearest neighbour classifier (with K=2) produced the decision boundaries in figure 5a. As you can see they are piecewise-linear which may suggest over-fitting; linear boundaries in the same general direction would classify the test data equivalently (see figure 5b) and may generalise better to future data.

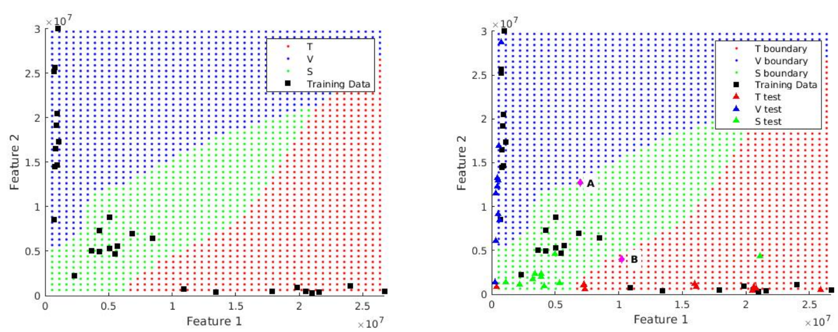
In figure 5b, you can see how the classifier performs on the test data. Generally, the classifier performs well. This may be influenced by the choice of using only 2 features. By avoiding less informative features it ensures the nearest neighbour classifier does not consider them equally to the informative features when calculating its decision boundaries. In addition, with fewer features, the classifier is less susceptible to over-fitting associated with the curse of dimensionality.

The test images which were incorrectly classified are shown in figure 4. The letter T in figure 4a, has a right-diagonal component, which means it does not give a high value for feature 1, corresponding to the horizontal line typical of the letter T. In addition, the right-diagonal is tilted at a small angle, meaning it is not characteristic of the letter V either, which has diagonal components closer to the vertical. Therefore it has very low feature values, as seen by the red triangle in the bottom left of figure 5b. The letter V in figure 4b was also classified incorrectly due to its very steep diagonal lines. This forms lines in the Fourier space close to the horizontal which is missed by the sector region of feature 2, explaining its position in the bottom left of figure 5b (blue triangle). In order to classify this character correctly, it would mean increasing the sector angle which would result in high values for the letter T which has a line in the horizontal direction of the Fourier domain. This trade-off results in mis-classification of particularly steep V characters and consequently highlights an area in which I could improve my features. Finally, the S character in figure 4c has a significant horizontal component which causes a large value for feature 1. Thus, the character is classified as the letter T (see far right of figure 5b).



(a) (b) (c)

Figure 4: Incorrectly classified characters from my test data set



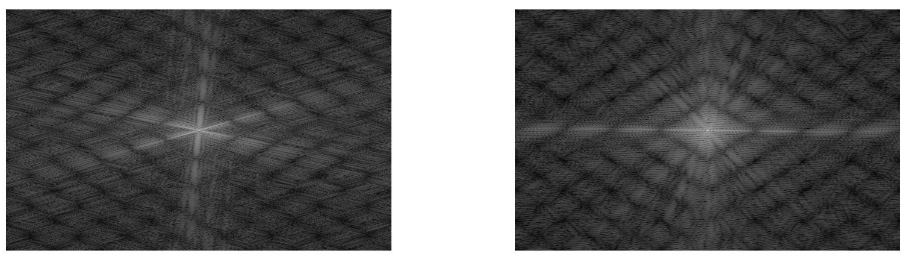
(a) (b)

Figure 5: a) A scatter plot of the feature values of the training data, showing the decision boundaries for the nearest neighbour classifier. b) A scatter plot showing the test (traingles) and training (squares) data plotted with the decision boundaries of the nearest neighbour classifier. The feature values of the letters A and B can also be seen.

5 Classification of letters A and B

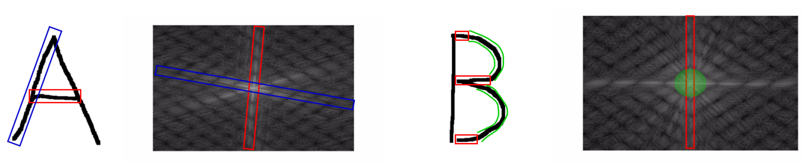
The classification of letters A and B are also shown in figure 5b. The letter A is classified as the letter S. The letter has a horizontal component which is captured by feature 1 to characterise the letter T; however the letter also has a right diagonal component which is captured by feature 2 to characterise the letter V (see figure 6a and 7a). Hence both feature values are moderately high and so the letter is classified as the letter S since its decision area occupies the central region of the feature space.

The letter B also has significant components of each feature. Concretely, the character has horizontal components, characteristic of the letter T, as shown by the faint vertical lines in the Fourier space (see figure 6b and 7b). It also curved edges, similar to that of the letter S. These attributes are captured by the combination of features 1 and 2. Its position in figure 5b is representative of this, positioned very close to the boundary between T and S (see figure 5b).



(a) (b)

Figure 6: Fourier transforms of the images for a) A and b) B



(a) (b)

Figure 7: Components in the letters A and B captured by features 1 and 2, coloured by their presence in the letters T (red), V (blue) and S (green) along with their associated Fourier transforms. [2]

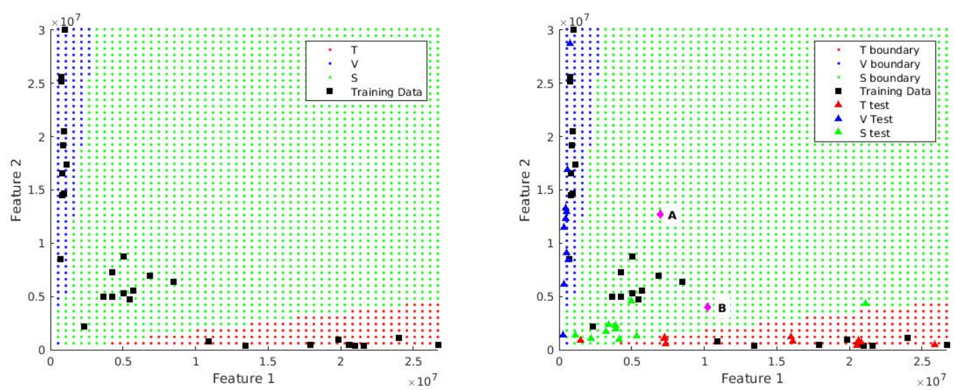
[2] In figure 7b, the green circle represents magnitudes in all directions. Although the circle has a set radius, this is not representative of the Fourier transform which has magnitudes in the higher frequencies which cannot be seen clearly due to their low intensity.

6 Alternative Classifier

Due to the relatively strong negative correlation between my features, it would not be appropriate to apply a naive Bayes classifier as this assumes independent features. Also, because of the limited amount of training data, it may not be appropriate to use a decision tree classifier which is very sensitive to slight variations in the training data. On the other hand, using a maximum likelihood classifier accounts for trends in the data, using mean and variance parameters (assuming a normal distribution) and so may be able to capture the tendency of the T training features to spread across the x axis, and of the V training features to spread across the y axis. In addition, because it is a probabilistic model, it could provide class membership probabilities. This could be particularly useful for points with low feature values (i.e. bottom left of figure 5a), where the classes are less separable and classification is more difficult.

Applying a maximum likelihood classifier produced the decision boundaries in figure 8a. For each class the mean vector and covariance matrix was estimated and then used to construct to a normal distribution. The decision boundaries are then formed by the points (x, y) for which $P((x, y)|T) = P((x, y)|V)$, $P((x, y)|T) = P((x, y)|S)$ or $P((x, y)|V) = P((x, y)|S)$.

As predicted above, the covariance matrix estimated for the feature values of T show very low variance for feature 2 (0.0007 after normalisation) and equivalently feature 1 for the letter V (0.0002 after normalisation). This results in very thin Gaussian curves and thin regions of figure 8a assigned to the two classes. This leads to the interpretation that everything not characteristic of a T or V, is classified as an S, agreeing with the logic used to choose my features originally, and explain the classifier's classification of the characters A and B (see figure 8b). Accounting for these small variances may allow the model to generalise well to future data. This is evidenced by its performance on the test data in figure 8b. The maximum likelihood classifier classified more of the test instances correctly. In particular the S in figure 4c was previously considered an outlier by the nearest neighbour classifier, however is correctly classified by the maximum likelihood classifier. Despite this, both classifiers incorrectly classified the T and V characters in figure 4. Visually, these characters are clearly identifiable and thus highlights a need for more features or training data.



(a) (b)

Figure 8