Sacha Raman
1288048

ASSIGNMENT 3 INVESTIGATION REPORT

To evaluate and determine the effectiveness of my image descriptor, I have decided to compare various car, plane and train images through the image-similarity function. This report will contain an outline of my experiment followed by the results obtained from conducting it. I will then analyse the results and will conclude the report with a summary of my investigation.

EXPERIMENT OUTLINE

I will start with a dataset of 15 images, containing 5 images of each of the three labels: car, train and plane. Then I will compare each image with all other images in the dataset through the image-similarity function and record my results. The image-similarity function compares the image-descriptor output for two images and returns a value between 0 – 1. The value shows the level of similarity between the two images, with a value of 1 meaning the two images are identical. The further it deviates from 1, the greater the dissimilarity in the two images. The image-descriptor function returns a vector of length 24 containing the normalized edge-direction, magnitude and intensity histograms. After I have compared all the images in my dataset, I will analyse the data in three ways:

- Separating the data by comparisons between labels then calculating the average similarity values for each set.
- Ranking the data in order of similarity based on the output of the image-similarity function (top 5).
- Organising the data by their similarity value.

RESULTS

The 15 images chosen for my dataset are images 5-9 for each type car, plane and train. After comparing all images in my dataset with each other, these are the similarity results I have obtained (rounded to 3dp):

Car & Car Comparisons
car5 & car6: 0.876
car5 & car7: 0.822
car5 & car8: 0.841
car5 & car9: 0.877
car6 & car7: 0.814
car6 & car8: 0.858
car6 & car9: 0.847
car7 & car8: 0.798
car7 & car9: 0.882
car8 & car9: 0.847

Plane & Plane Comparisons
plane5 & plane6: 0.467
plane5 & plane7: 0.762
plane5 & plane8: 0.756
plane5 & plane9: 0.791
plane6 & plane7: 0.602
plane6 & plane8: 0.691
plane6 & plane9: 0.542
plane7 & plane8: 0.807
plane7 & plane9: 0.783
plane8 & plane9: 0.777

Train & Train Comparisons
train5 & train6: 0.778
train5 & train7: 0.873
train5 & train8: 0.799
train5 & train9: 0.791
train6 & train7: 0.740
train6 & train8: 0.852
train6 & train9: 0.843
train7 & train8: 0.840
train7 & train9: 0.857
train8 & train9: 0.896

Car & Plane Comparisons
car5 & plane5: 0.857
car5 & plane6: 0.465
car5 & plane7: 0.777
car5 & plane8: 0.719
car5 & plane9: 0.741
car6 & plane5: 0.830
car5 & plane6: 0.538
car6 & plane7: 0.751
car6 & plane8: 0.790
car6 & plane9: 0.762
car7 & plane5: 0.736
car7 & plane6: 0.470
car7 & plane7: 0.694
car7 & plane8: 0.669
car7 & plane9: 0.651
car8 & plane5: 0.855
car8 & plane6: 0.530
car8 & plane7: 0.794
car8 & plane8: 0.794
car8 & plane9: 0.828
car9 & plane5: 0.794
car9 & plane6: 0.452
car9 & plane7: 0.734
car9 & plane8: 0.682
car9 & plane9: 0.687

Train & Plane Comparisons
train5 & plane5: 0.678
train5 & plane6: 0.538
train5 & plane7: 0.767
train5 & plane8: 0.668
train5 & plane9: 0.651
train6 & plane5: 0.729
train5 & plane6: 0.429
train6 & plane7: 0.615
train6 & plane8: 0.588
train6 & plane9: 0.549
train7 & plane5: 0.767
train7 & plane6: 0.552
train7 & plane7: 0.838
train7 & plane8: 0.761
train7 & plane9: 0.753
train8 & plane5: 0.818
train8 & plane6: 0.499
train8 & plane7: 0.756
train8 & plane8: 0.692
train8 & plane9: 0.669
train9 & plane5: 0.859
train9 & plane6: 0.488
train9 & plane7: 0.746
train9 & plane8: 0.720
train9 & plane9: 0.691

Train & Car Comparisons
train5 & car5: 0.802
train5 & car6: 0.753
train5 & car7: 0.862
train5 & car8: 0.753
train5 & car9: 0.812
train6 & car5: 0.794
train5 & car6: 0.765
train6 & car7: 0.791
train6 & car8: 0.710
train6 & car9: 0.815
train7 & car5: 0.857
train7 & car6: 0.815
train7 & car7: 0.819
train7 & car8: 0.851
train7 & car9: 0.846
train8 & car5: 0.874
train8 & car6: 0.827
train8 & car7: 0.799
train8 & car8: 0.816
train8 & car9: 0.847
train9 & car5: 0.908
train9 & car6: 0.838
train9 & car7: 0.824
train9 & car8: 0.852
train9 & car9: 0.869

ANALYSIS

My first method of analysing the data involves calculating the average of each of the six sets of data then using a bar graph to visually demonstrate the averages.

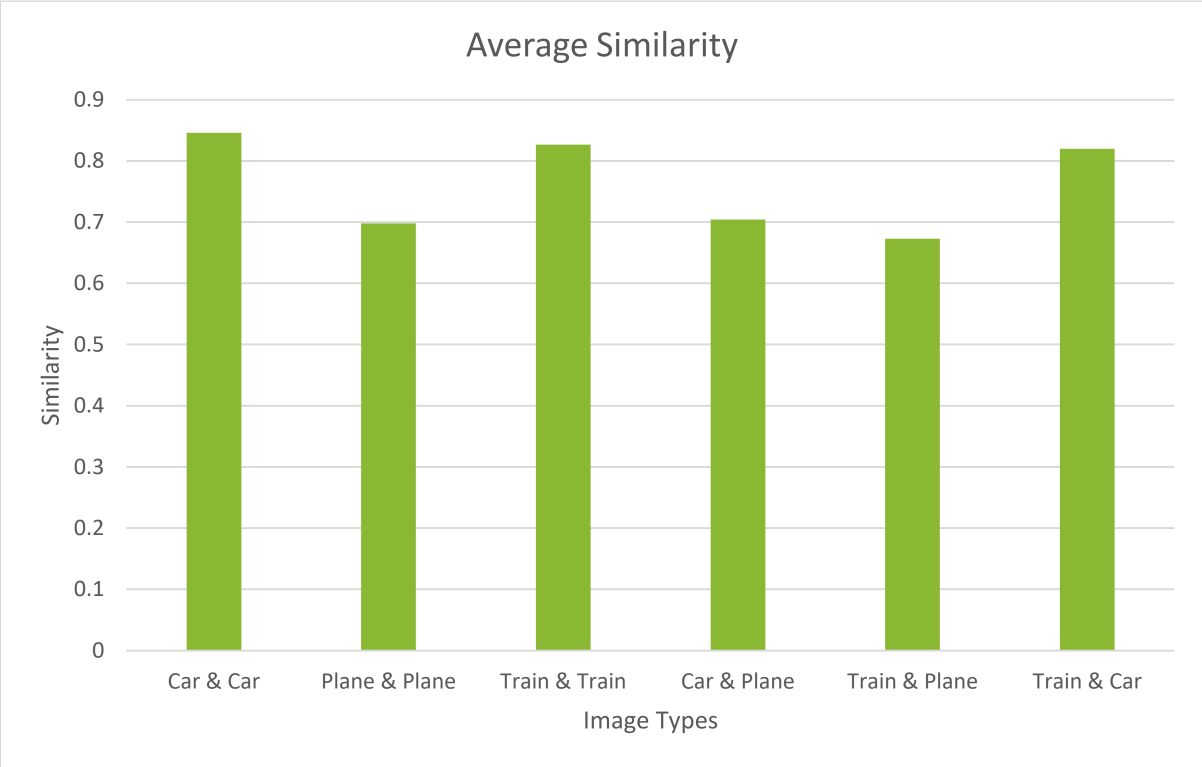

Figure 1: The average similarity value between 0 – 1 for each comparison group.

From this graph we can see that the average value outputted from the image-similarity function appears to be around what we would expect from two car images or two train images as they both have a common factor to them. Both of these averages are over 0.8, showing a significant similarity between the two images under comparison. On the other hand, the average image-similarity value outputted for two plane images is slightly surprising and not quite what would be expected. The average is just shy of 0.7 which albeit high, is not quite as high as the values between two car or two train images. This could potentially be due to the different possible angles for taking a picture of a plane in comparison to those of a train or car. With the train and car images, they are all taken from an angle, generally showing the front and a side of the vehicle. In comparison, the some of the plane images also show the underside of the plane when the picture is taken of the plane in mid-flight. Another feature to take notice of is that with the car images, the cars are always on the ground on some sort of road or driveway and the trains are always on tracks. Whereas with the planes some are on the ground and others are in the air. This environmental difference would probably have an impact when comparing similarities between two images.

The last three bars in this graph show the average value outputted from the image-similarity function between a car and a plane, a train and a plane, and a train and a car. The average value between a car and a plane is actually slightly higher than the average value between two planes. Furthermore, the average value between a train and a car is well above the value between two planes and is almost as high as the value for two trains. I found this mildly concerning as it suggests that the image-descriptor is not very accurate at all. The average value between the train and the plane slightly restores my confidence in the image-descriptor as it is the lowest value on this graph (0.673) however in my opinion it is still high for two very different vehicle types.

The next thing I will do is rank the top 5 image-similarity values:

Top 5 Image-Similarity Values

1. train9 & car5: 0.908

2. train8 & train9: 0.896

3. car7 & car9: 0.882

4. car5 & car9: 0.877

5. car5 & car6: 0.876

Figure 2: A chart showing the top five similarity values and the images they correspond to.

The most interesting thing about the top 5 values, is that the first value is not even between two images of the same label, it is between a train and a car, and while it can be observed from Figure 1 that the average value between a train and a car image is quite high, I was still not expecting the highest image-similarity value to be between images two images of different vehicle types.

I then expected the second highest similarity value to be between two car images as the image-similarity between two cars had the highest average value, however, it is actually between two train images. This isn't completely unexpected though as the average value between two trains was the second highest according to Figure 1.

The next three top similarity values are between two images of cars which makes sense. Initially I was confused with the two cars places third (car7 and car9) due to their contrasting colours. The only explanation I could come up for this high similarity value was that in the car7 image, the background is quite light which is similar to the car in the car9 image. Likewise, the background in the car9 image is similar to the colour of the car in the image car7. Unsurprisingly there are no plane images in the top 5, however this was to be expected based on the averages calculated in Figure 1.

Finally, I will order the data based on their similarity value:

Image-Similarity Frequency Table

	0.4 – 0.5	**0.5 – 0.6**	**0.6 – 0.7**	**0.7 – 0.8**	**0.8 – 0.9**	**0.9 – 0.1**	**Total**
Car & Car	0	0	0	1	9	0	10
Plane & Plane	1	1	2	5	1	0	10
Train & Train	0	0	0	4	6	0	10
Car & Plane	3	2	5	11	4	0	25
Train & Plane	3	4	7	8	3	0	25
Train & Car	0	0	0	7	17	1	25
Total	7	7	14	36	40	1	105

Figure 3: A frequency table showing the counts of image-similarity values for each range for each comparison group.

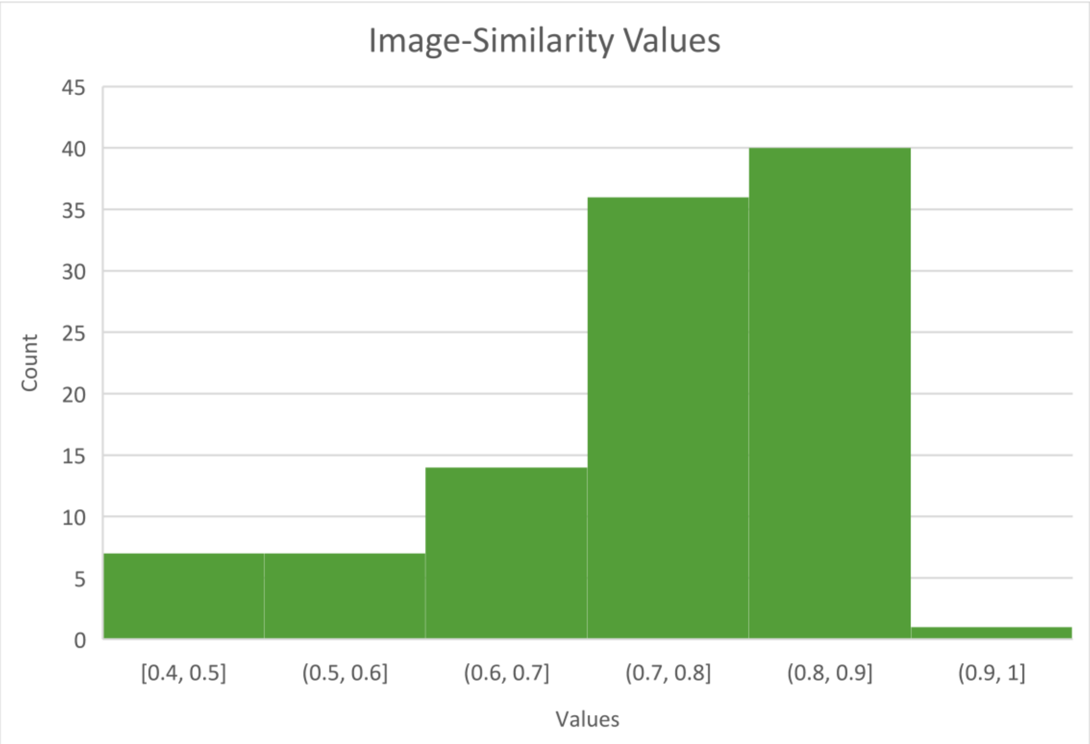

Figure 4: A histogram displaying the data from the frequency table.

As expected, car & car comparisons had quite high results, with 9 out of the 10 comparisons being in the 0.8 – 0.9 range. Likewise, with train & train comparisons, over half of the image-similarity values placed between 0.8 and 0.9. However, with the plane & plane comparisons, we can see that 4 of the 10 image-similarity values are below 0.7 and only one comparison results in a value over 0.8.

The results for car & plane and train & plane are quite spread out whereas the results for train & car have a very small range, mainly spanning between 0.7 – 0.9. 17 out of 25 of the values placed between 0.8 – 0.9 for this category which is significantly high for images of different vehicle types.

CONCLUSION

Overall I am quite surprised by the results obtained from my investigation and I am rather sceptical on the accuracy of the image-descriptor based on the values produced by the image-similarity function. The results seem fairly normal for two car images or two train images however the values for two images containing different vehicle types are rather high making the image-descriptor quite unreliable. For this investigation my dataset of images was quite small, containing only 5 images of each vehicle type (a total of 15 images). I believe this would have had an effect on the overall results. For a better reflection of the image-description function, a larger dataset would be ideal as this would reduce the chance of data being skewed by outliers.

It would be interesting to see if increasing the number of bins from 8 to 16 or 32 for each component in the image-descriptor would make any difference or perhaps increase the accuracy slightly? Or pre-processing all the images first, by applying a filter over all images such as a blur filter. Also, another thing of interest would be the image-descriptor itself and what components have the greatest effect when comparing the similarity of images and if any of these can be enhanced to produce more accurate results. In this experiment, I only used images containing vehicles, however, I am curious what the results would be on images containing things such as animals or plants.