Name: Tianxiu Ma

Student number: 230277302

Assignment number: Assignment 1

module code: ECS797U/ECS797P

Task 1: Explain the input and the output of the K-means; what are the input and output dimensions? what is the use of the dictionary? [3 marks]

a. **The input** to this process consists of a subset of feature vectors, known as descriptors, derived from the training dataset. These descriptors serve to encapsulate distinctive keypoints within an image. Specifically, the input is organized into a two-dimensional array, where each row corresponds to a unique descriptor. The array **dimensions are (6000, 128),** indicating that there are 6000 such descriptors, each characterized by 128 features.

The n_clusters parameter, set to 500, specifies the desired number of clusters, or "words," to be formed from the input descriptors, effectively creating a visual dictionary. The algorithm parameter, set to 'elkan', indicates the use of the Elkan variation of the K-means algorithm, which is typically more efficient for processing dense datasets, albeit at the expense of higher memory usage. Additionally, the random_state parameter ensures the consistency of results across different runs by initializing the algorithm's random number generator with a predetermined seed.

b. The output is the resulting dictionary (cluster centers) with dimensions (K, 128), where K is the number of words in the dictionary. Each cluster center (visual word) can be thought of as a representative of a group of similar descriptors. In this case, **the dimensions of output are (500,128).**

c. **Use of the dictionary:** The dictionary crafted from cluster centers, or "visual words," provides a succinct representation of the visual features within the dataset, facilitating tasks such as image classification. This is achieved through the quantization process, where descriptors from both training and test images are assigned to their closest visual word in the dictionary. Such quantization transforms continuous feature vectors into discrete visual words, enabling a more structured and interpretable approach to image analysis.

Task 6: Evaluate both classifiers on the test set (in terms of accuracy and confusion matrices) and discuss results. [3 marks]

The accuracy and confusion matrices results obtained on the test set for the two classifiers are as follows:

```
Euclidean Distance  Accuracy: 0.74
Euclidean Distance Confusion Matrix:
 [[18  1  0  0  1]
 [ 1 19  0  0  0]
 [ 1  2 13  3  1]
 [ 0  0  1 19  0]
 [ 1  1 10  3  5]]
Histogram Intersection Accuracy: 0.16
Histogram Intersection Confusion Matrix:
 [[ 2  0  4  0 14]
 [ 9  0  0  0 11]
 [ 7  0  0  0 13]
 [ 3  0  0  0 17]
 [ 5  0  0  1 14]]
```

Euclidean Distance Classifier:

a. **Accuracy:** The accuracy of the Euclidean distance classifier is 0.74. It means that 74% of the dataset was correctly classified, indicating a relatively high level of performance for this classifier.

b. **Confusion Matrix:**

The numbers on the diagonal (18, 19, 13, 19, 5) indicate how many times each class was accurately predicted. Misclassifications are represented by the off-diagonal numbers. This demonstrates that the classifier generally performs well for most classes. However, there's notable confusion primarily among classes 3, 4, and 5. This suggests that there are certain similarities between these classes that the Euclidean distance metric has difficulty differentiating.

Histogram Intersection Classifier:

a. **Accuracy:** The accuracy is significantly lower at 16%, suggesting this classifier performs poorly on the test set.

b. **Confusion Matrix:**

The correct predictions are notably fewer (2, 0, 0, 0, 14), with a large number of instances being misclassified into the last class, suggesting that the histogram intersection method may not be effective for this particular dataset.

Discussion of results:

Comparing the two classifiers, it's evident that the method based on Euclidean distance far outperforms the histogram intersection approach in this scenario. This could be due to the Euclidean distance's better capability to capture differences between samples when dealing with such data. On the other hand, the histogram intersection method might be biased towards predicting a particular class due to its way of calculating similarity, which does not suit the distribution or type of the dataset used for testing. Furthermore, the extremely low accuracy of histogram intersection suggests that this method is highly sensitive to data preprocessing and feature selection, and it might perform better under different parameter settings or data representations. Overall, for this specific application, traditional distance measures, like Euclidean distance, may be more reliable than histogram intersection, especially when dealing with feature vectors of high-dimensional data, where geometric properties are crucial for classification.

Task 7: Repeat steps 1-6 using a very small dictionary size (eg. 5). Compute the accuracy and confusion matrices and discuss the drop in performance [2 marks]

The accuracy and confusion matrices of small dictionary size:

```
Small Dictionary Size Euclidean Distance Accuracy: 0.4
Small Dictionary Size Euclidean Distance Confusion Matrix:
 [[ 7  0 12  0  1]
 [ 5 11  4  0  0]
 [ 6  0 14  0  0]
 [11  0  7  2  0]
 [ 9  0  5  0  6]]
Small Dictionary Size Histogram Intersection Accuracy: 0.19
Small Dictionary Size Histogram Intersection Confusion Matrix:
 [[ 1  0  3 10  6]
 [ 3  0  4  6  7]
 [ 1  0  5  4 10]
 [ 1  2  3  1 13]
 [ 1  1  3  3 12]]
```

Euclidean Distance:

The utilization of a larger dictionary facilitates a more nuanced and detailed representation of the visual features present in images, thereby enhancing the accuracy of classifications. A diminished dictionary size constrains this capability, resulting in less distinctive visual words and a greater degree of ambiguity in classifying tasks. This reduction in performance is particularly notable in the case of the Euclidean distance classifier, which depends significantly on the detail within the feature space to achieve accuracy. Consequently, a smaller dictionary size impairs the classifier's proficiency in discerning and differentiating the subtle characteristics of various classes, culminating in elevated rates of misclassification.

Histogram Intersection:

Although there is a slight increase in accuracy from 0.16 to 0.19 with a smaller dictionary size, the performance remains substantially low. The slight improvement in accuracy might be due to the reduced complexity making it easier to match histograms. The overall method still struggles to accurately classify the dataset and the method's inherent limitations in capturing the relevant features for classification remain evident.

Task 8: Evaluate SVC classifier on the test set (in terms of accuracy and confusion matrices) and compare to the KNN classifier results.

For each class show some images that are correctly classified and some images that are incorrectly classified.

Can you explain some of the failures? [3 marks]

The accuracy and confusion matrices of the SVC Classifier:

```
SVC Classifier Accuracy: 0.84
SVC Classifier Confusion Matrix:
 [[19  0  1  0  0]
 [ 0 19  1  0  0]
 [ 3  0 15  1  1]
 [ 0  0  3 17  0]
 [ 1  0  5  0 14]]
```

We can see that, the accuracy of SVC classifier is 0.84, which is higher than KNN classifier results. The confusion matrix of SVC classifier indicates strong performance across most classes with few misclassifications. The SVC is better at capturing the nuances of the feature space and making more accurate class distinctions. And its robustness in handling overlaps between classes. So the SVC classifier outperforms the KNN classifier.

Classification Failures explanation:

Here are some reasons:

a. **Insufficient or Inaccurate Feature Extraction**: SVC relies on the features of input data for classification. If the features used for classification do not sufficiently differentiate between categories, the model may fail to correctly identify specific objects. For instance, if feature extraction fails to capture the unique visual differences between a "dog's profile" and a "car," the model might confuse these two categories.

b. **Similarity Between Categories**: Some objects may share visual features that make it difficult for classifiers to distinguish between them. For example, the profile of a dog and the silhouette of certain vehicles might appear similar under some feature extraction methods; airplanes and vehicles might share similarities in size, shape, or texture, leading to classification errors.

c. **Limited Feature Coverage**: The features used might not cover all the information crucial for classification. For example, a keyboard might be mistakenly identified as an airplane due to its rectangular shape and lines.

The following measures can be taken to improve performance:

a. **Increase Sample Diversity**: Ensuring the training data covers a diversity of each category, including different angles, lighting conditions, and backgrounds, can help improve the model's generalization capability.

b. **Consider Model Ensemble**: Combining the predictions of multiple models may reduce classification errors caused by the bias of a single model.

c. **Use data augmentation techniques**, such as rotation, scaling, cropping of images, etc., to increase the diversity of samples during model training and improve the model's generalization capability.