CSCI 574 Hw#5 : Object Recognition using Bag of Features

Name: Manan Vyas
USC-ID : 7483-8632-00
Email: mvyas@usc.edu

Task: To train a Bag of Features recognizer using the set of training images provided and then recognize the trained objects from unknown images using a vote based K-Nearest Neighbour classifier.

Code implemented in: C++, OpenCV (Build in CodeBlocks|MinGW)

Discussing in details the results obtained for the following combination of input parameters;

```
int const kComponents = 30 ;

int const kClusters = 100 ;

int const kNearestNeighbors = 10;
```

- Training: To generate the codewords during the training phase, the following steps were implemented;

  1. SIFT Features : Local SIFT Features were determined using the SIFT::operator()[1] function in OpenCV. Once these 128 dimension features are obtained, their complexity was to be reduced by doing a Principal Component Analysis on these features. For this purpose, a fixed number of initial eigen vectors that correspond to the principal directions of these features were taken into account. I experimented with a varying number of these initial number of prinicipal direction (here taken to be 30) in addition to the recommended 20 components given in the problem. The dot product of the SIFT features and the principal direction gives us the PCA-SIFT features.

  2. Generating training codewords: Using a K-means clustering algorithm[2] the calculated PCA-SIFT features were clustered in a certain number of cluster centers. The effect of choice of number of clusters on the recognizer was also studied in addition to using the recommended number of 100 given in the problem.

  3. Feature Vector: Finally a feature vector set for a training images was generated by creating a histogram of the codewords obtained from step2 above. These can now be used by the recognizer to make a decision. A few of the histograms are plotted in the results section below.

- Recognizer

  1. N-Nearest Neighbour Recognizer: A N-Nearest neighbour recognizer was implemented. Herein, PCA-SIFT features of the testing image

were calculated. Then the distance each feature from the center was computed. The following code snippet illustrates my implementation of determining the distance of each feature from the center using the conventional distance formula ;

```
for(int i=0; i<feature.rows; i++)
{
	for(int j=0; j<(kClusters); j++)
	{
		for(int k=0; k<(kComponents); k++)
		{
			distance[i][j] += pow(feature.at<double>(i,k) -
			centers.at<float>(j,k), 2.0);
		}
		distance[i][j] = sqrt(distance[i][j]);
	}
}
```

2. Histogram Generation: Once, the distances of features from the centers were calculated, the histogram of the ascending arranged distances was created. This histogram will now be used to used to recognize the object by selecting the image whose training and testing histograms have minimum Euclidean difference.

3. Voting based on nearest neighbours: A voting based decision system was created using the ascendingly sorted array of the minimum differences between the two histograms from step2. A specific number of nearest neighbours was used to make this voting decision and the effect of varying the nearest neighbour on recognizer was also studied. The following code snippet illustrated my implementation;

```
for(int i=0; i<(kNearestNeighbors); i++)
{
	if(sortedDifference.at<int>(i) >= 0 && sortedDifference.at<int>(i) < 20)
		vote[0] += diffBetweenHistograms[sortedDifference.at<int>(i)];

	else if(sortedDifference.at<int>(i) >= 20 && sortedDifference.at<int>(i) < 40)
		vote[1] += diffBetweenHistograms [sortedDifference.at<int>(i)];

	else if(sortedDifference.at<int>(i) >= 40 && sortedDifference.at<int>(i) < 60)
		vote[2] += diffBetweenHistograms [sortedDifference.at<int>(i)];

	else if(sortedDifference.at<int>(i) >= 60 && sortedDifference.at<int>(i) < 80)
		vote[3] += diffBetweenHistograms [sortedDifference.at<int>(i)];

	else
		vote[4] += diffBetweenHistograms [sortedDifference.at<int>(i)];
}
```

4. Decision : Based on the index position of the maximum number of votes in the vote array of 5 elements with index0->4 representing image set of {‘car’, ‘face’, ‘cougar’, ‘pizza’, ‘sunflower’} respectively, I select my recognized object.

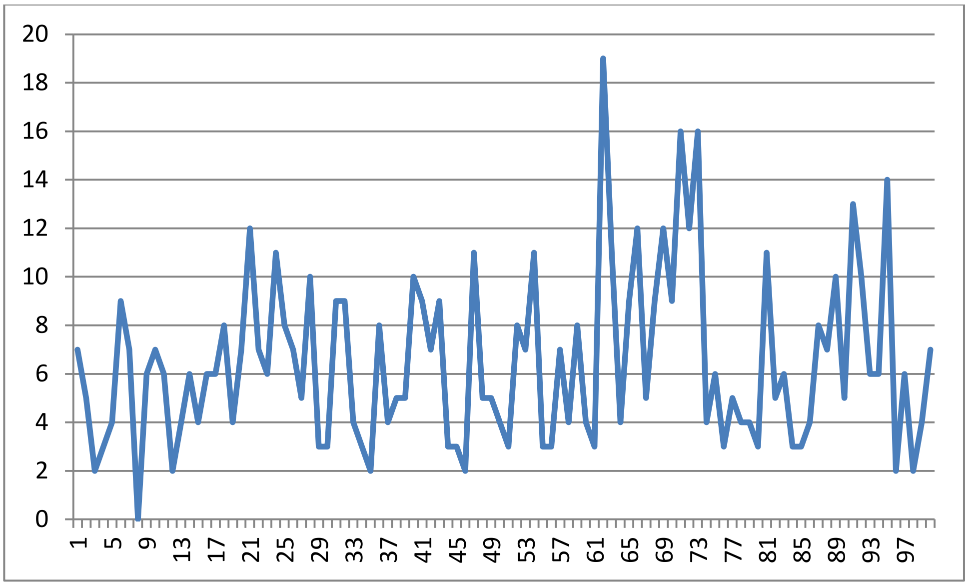

Fig1. Histogram of the feature vectors for the 100 clusters for car image in the training set

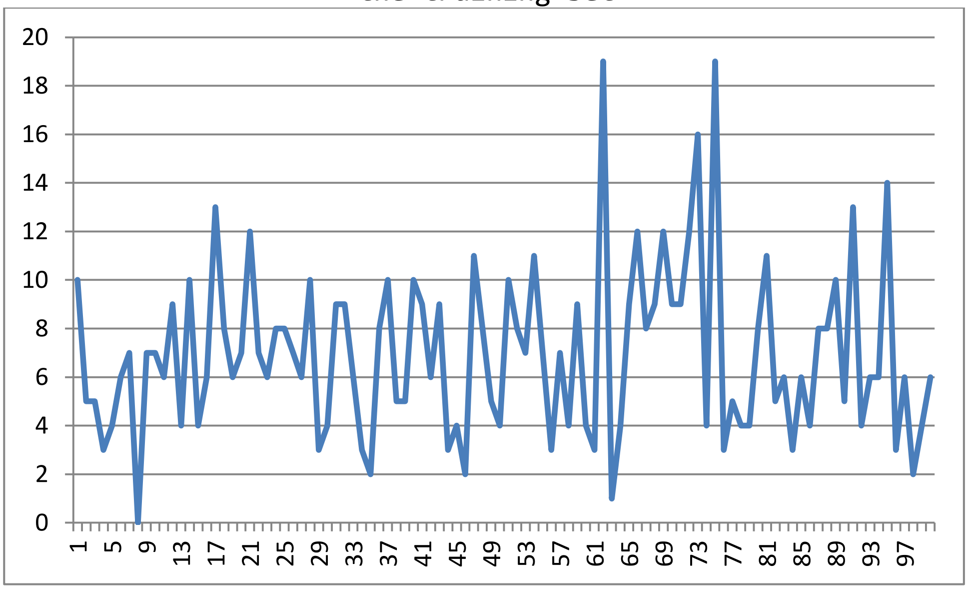

Fig2. Histogram of the feature vectors for the 100 clusters for car image in the testing set.

We can observe that the general trends for the car features are similar and it turns out that all of the testing images for car with the above parameters were identified correctly.

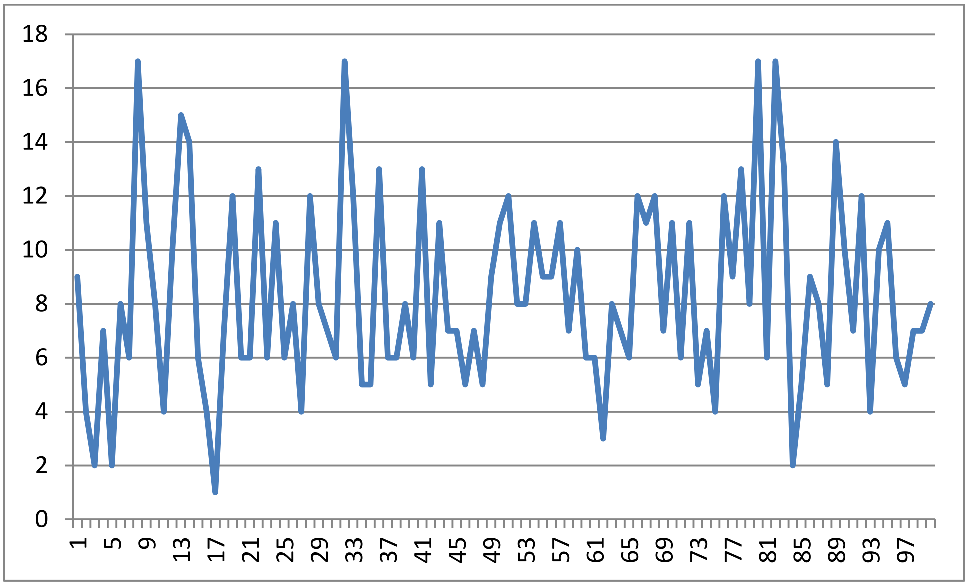

Fig3. Histogram of the feature vectors for the 100 clusters for cougar image in the training set

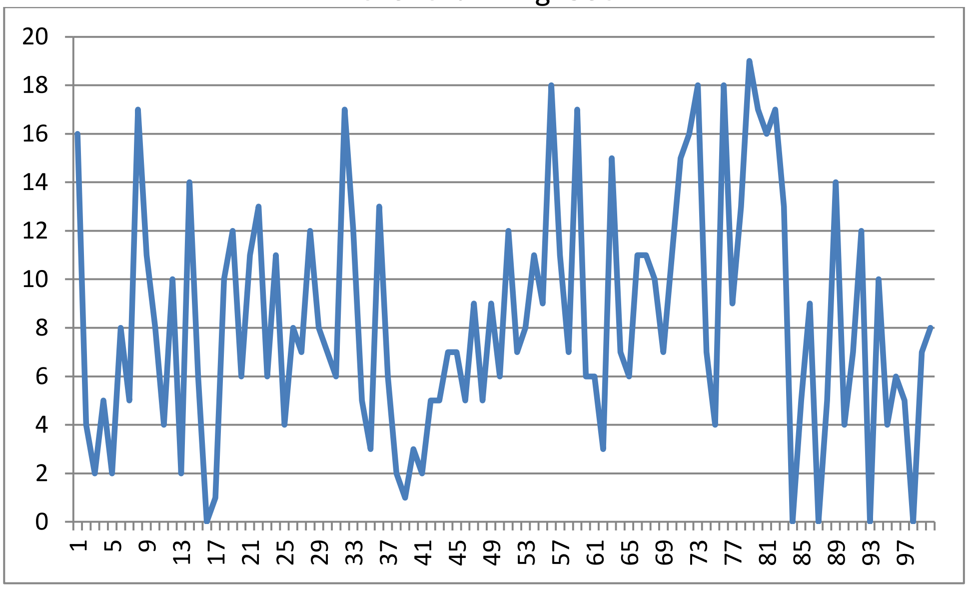

Fig4. Histogram of the feature vectors for the 100 clusters for cougar image in the testing set

We can observe that the general trends for the cougar features are not very coherent and it turns out that the recognition accuracy for cougar with the above parameters is lesser than the car and there are many mis-classifications of cougar with sunflower and some misclassification with face and pizza as well.

Here is how outputs were displayed on the command prompt;

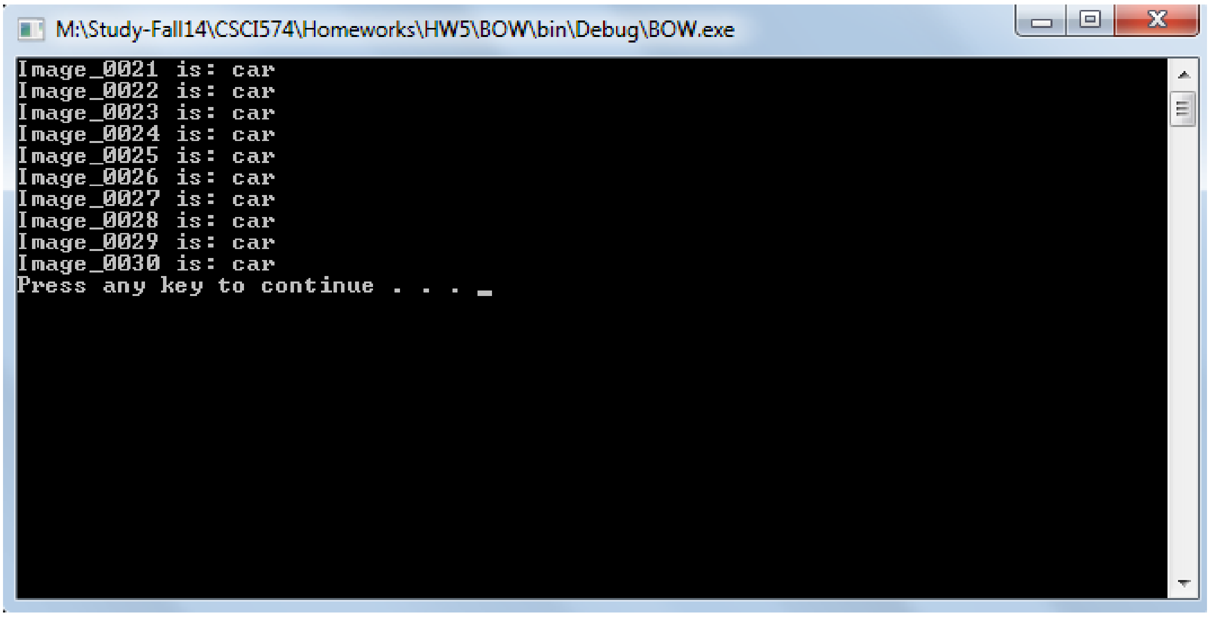

Fig5. Output for car testing images for the aforementioned parameters

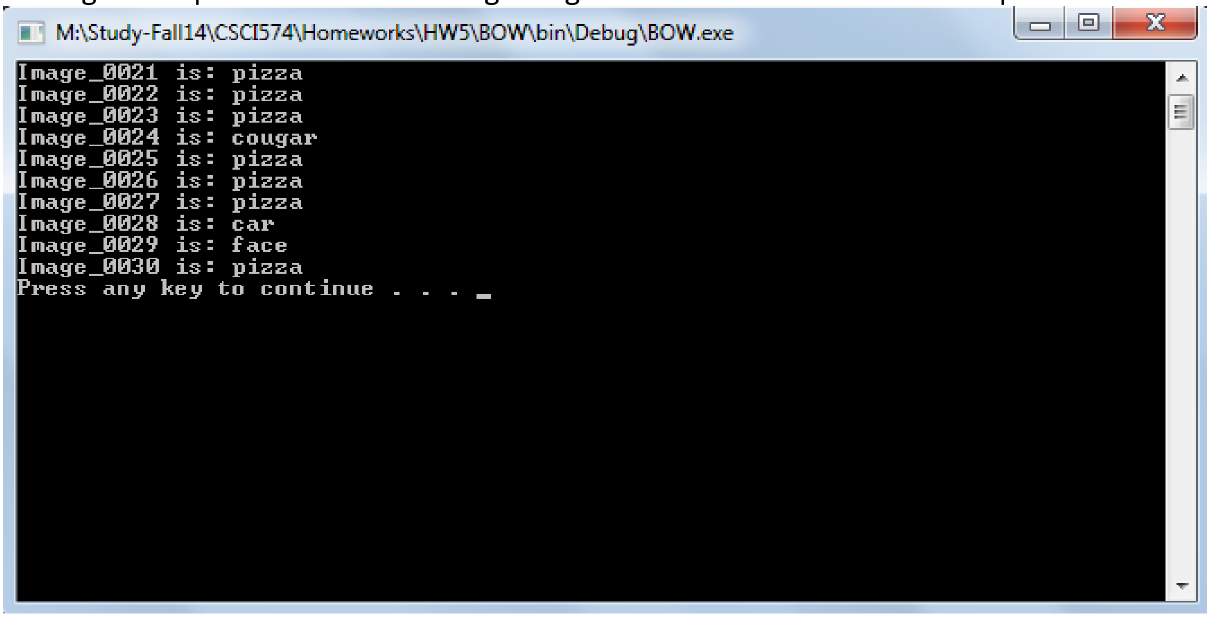

Fig6. Output for pizza testing images for the aforementioned parameters

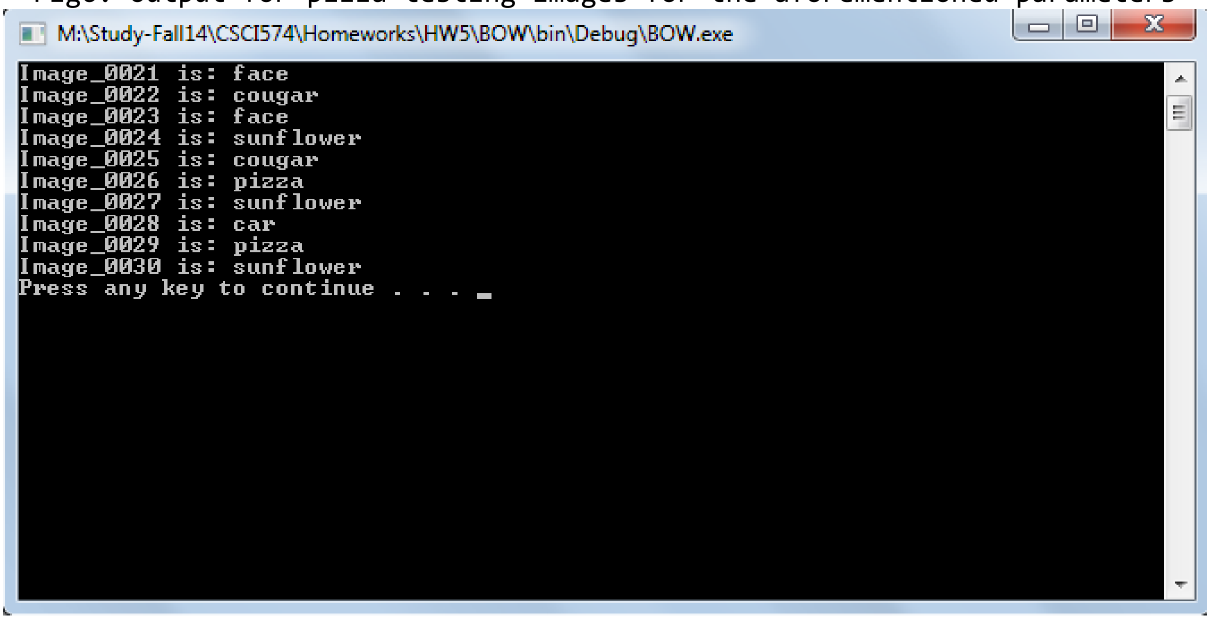

Fig7. Output for sunflower testing images for the aforementioned parameters

Results and Discussion:

The recognizer was run 10 times for different parameter combination.The KNN recognizer worked best for car and face followed by pizza, cougar and sunflower in that order. Face and car were identified correctly with accuracy>90% and in case of car it was 100% in most of the cases.Some mis-classification were observed in case of cougar and face but those were outliers around 15%. Pizza was also consistently identified at around 60-70% accuracy followed by cougar with 40-50% and sunflower recognition was least accurate with 30-40% accuracy. This may be because in many observations, sunflower-face clustering were close to each other so were cougar-sunflower, and face-sunflower and more often than not sunflower and cougar were mis-classified rather than the face. Similar is the case with face-cougar-pizza clustering wherein the colour variation and distribution of those 3 are similar and KMeans would cluster cougar-pizza-sunflower close to face rather than the other way round. This explains the high face recognition accuracy.

Effect of Variation of Parameters on the Recognizer:

Number of PCA components: PCA-based SIFT local descriptors are more distinctive, more robust to image deformations, and more compact than the standard SIFT representation. The main effect of this resulting into faster training and matching as the dimensionality of the problem is reduced. Comparing the results where the clusters are constant at 100 and nearest neighbours are constant at 10 but only principal components are varied from 20 to 30, it is found through the confusion matrix that higher number of principal components gave more recognition accuracy in the case of car (90% -> 100%) and sunflower (30% -> 40%) in this particular case. The cost of this being longer running time and complexity of the recognizer.

Number of K-Means Clusters: The level of detail within the given image decides the optimum level of clusters. An image with uniform components will be unnecessarily be split the uniform object into different clusters even though they belong to the same parent object because the KMeans overclassified due to higher number of clusters.The same gos other way around , the level of detail will be lost if an image with many objectsif KMeans is performed using low amount of clusters. In our case since the car occupies a bigger area within the image and we are not really interested in the background, lesser number of clusters may favour the car image but it also depends on what we have as the background. Face and cougar would be similar in clustering mechanism as the eyes, nose, ears etc. features of face are similar in number even thoughtheyr size and texture may differ. Similiar is the case with pizza and sunflower. Sunflower being hardest as essentially it has only 3 components the center, the petals and stem and ideally they should be clusters as that but since the number of clusters is high, every petal and pixel variation within the center will be clustered differently giving less accuracy.

Number of Nearest Neighbours for Recognizer: Larger value of nearest neighbours avoids misclassification and also avoids unnecessary extraneous or irrelevant features. However, as nearest neighbours were increased, the difference or the boundaries between the different votes became lesser and lesser and again misclassification began to occur as the difference between votes was so less that the recognizer no longer was able to sample those as different votes. This is particularly evident in car and face where as the nearest neighbours increased car image began to degrade in terms of recognition accuracy. So was the case with sunflower and face as well. A more robust method like SVM could also be used in place of KNN in our case.

Confusion matrix and error analysis for 4 such observations are given as follows;

Error Analysis: Following are the confusion matrices and accuracy statistics for four different combinations of input parameters;

- For the input parameters PCA-Components=20 ; K-Means Clusters = 125 and Nearest Neighbours = 15

	Car	Cougar	Face	Pizza	Sunflower
Car	1	0	0	0	0
Cougar	0.4	0.4	0.1	0.1	0
Face	0.3	0	0.7	0	0
Pizza	0.2	0.1	0	0.7	0
Sunflower	0.2	0	0	0.2	0.6

$$Accuracy = \frac{trace\ of\ the\ matrix}{(\sum all\ matrix\ elements)} [3]$$

$$Accuracy = 68\%$$

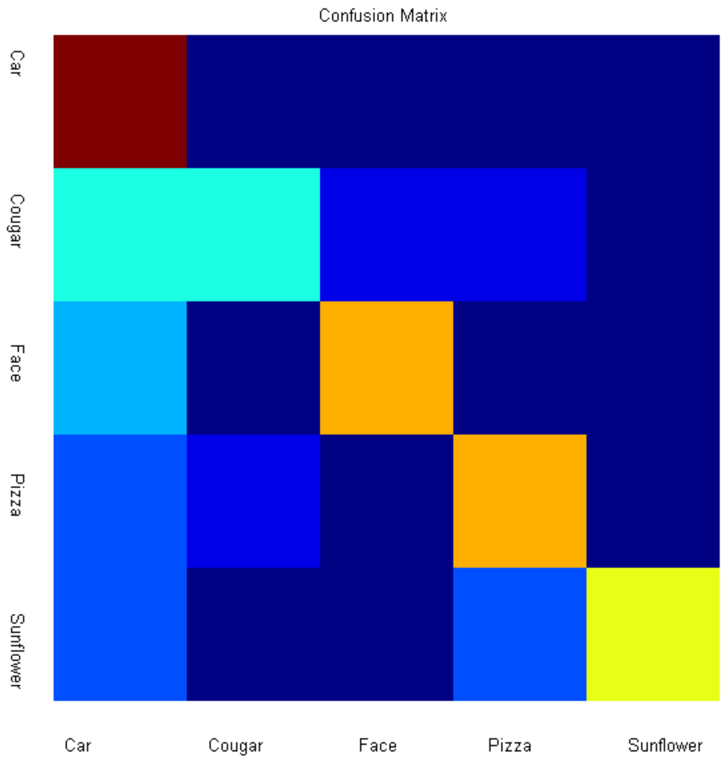

Fig8. Confusion Matrix for PCA-Components=20 ; K-Means Clusters = 125 and Nearest Neighbours = 15.

- For the input parameters PCA-Components=20 ; K-Means Clusters = 100 and Nearest Neighbours = 10.

	Car	Cougar	Face	Pizza	Sunflower
Car	0.9	0.1	0	0	0
Cougar	0.3	0.5	0	0.2	0
Face	0	0	1	0	0
Pizza	0.2	0.1	0	0.7	0
Sunflower	0.1	0.1	0.2	0.2	0.4

$$Accuracy = \frac{trace\ of\ the\ matrix}{(\sum all\ matrix\ elements)}$$

$$Accuracy = 70\%$$

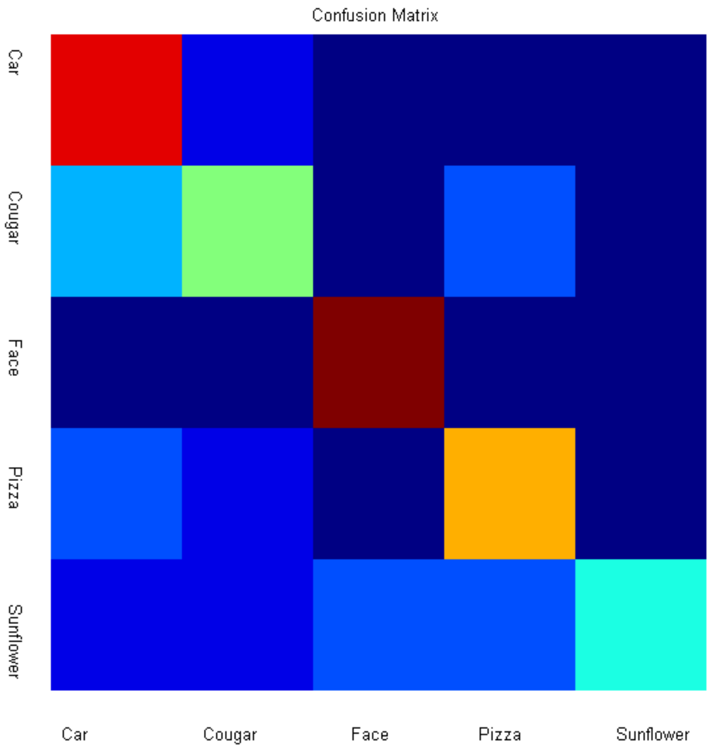

Fig9. Confusion Matrix for PCA-Components=20 ; K-Means Clusters = 100 and Nearest Neighbours = 10.

- For the input parameters PCA-Components=30 ; K-Means Clusters = 100 and Nearest Neighbours = 10.

	Car	Cougar	Face	Pizza	Sunflower
Car	1	0	0	0	0
Cougar	0.2	0.5	0.1	0.2	0
Face	0	0	1	0	0
Pizza	0.2	0.1	0	0.7	0
Sunflower	0.1	0.1	0.2	0.2	0.4

$$Accuracy = \frac{trace\ of\ the\ matrix}{(\sum all\ matrix\ elements)}$$

$$Accuracy = 73.46\%$$

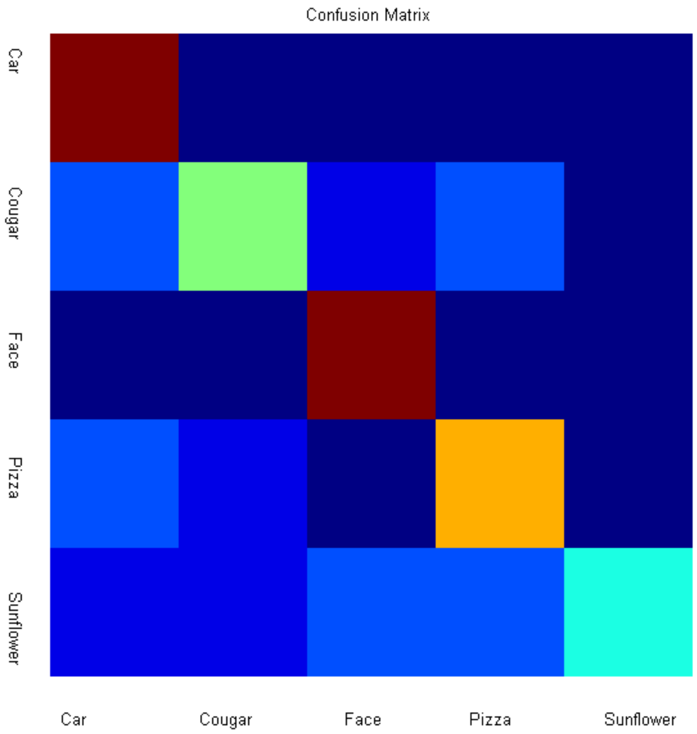

Fig10. Confusion Matrix for PCA-Components=30 ; K-Means Clusters = 100 and Nearest Neighbours = 10.

- For the input parameters PCA-Components=50 ; K-Means Clusters = 150 and Nearest Neighbours = 30.

  The numerical confusion matrix is calculated as;

	Car	Cougar	Face	Pizza	Sunflower
Car	0.9	0.1	0	0	0
Cougar	0.5	0.5	0	0	0
Face	0.4	0	0.6	0	0
Pizza	0.2	0.1	0.1	0.6	0
Sunflower	0.4	0.2	0.1	0	0.3

$$Accuracy = \frac{trace\ of\ the\ matrix}{(\sum all\ matrix\ elements)}$$

$$Accuracy = 58\%$$

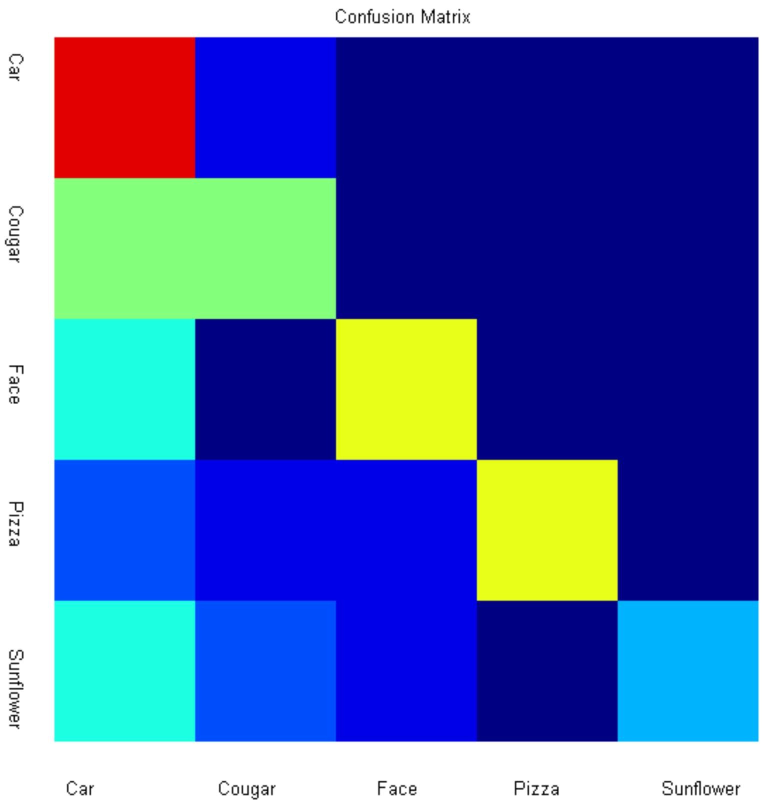

Fig11. Confusion Matrix for PCA-Components=50 ; K-Means Clusters = 150 and Nearest Neighbours = 30.

Similarly accuracy for all the 10 variations in parameters was calculated and average accuracy for the recognizer turned out be 71.44%

References:

[1]<URL>http://docs.opencv.org/modules/nonfree/doc/feature_detection.html

[2]<URL>< http://docs.opencv.org/modules/core/doc/clustering.html>

[3]<URL><www2.cs.uregina.ca/~dbd/cs831/notes/confusion_matrix/confusion_matrix.html>