# ECE 5554 : Computer Vision - Homework 5

Raghunath V P

November 20, 2017

## Problem A - Color histogram and k-nearest neighbor (kNN) classifier

**a.** In this problem we take a naïve approach to image classification and assume that classification can be made on the basis of presence/absence of certain hues in the image. For this classification task we have 8 classes. To achieve the objective, we calculate the histogram of color intensities in each of the 3 channels for each image. This is done by reshaping the 2D color intensity matrix in each channel separately into a 1 D vector and calculating the histogram using the hist function. In my implementation, I have selected the number of bins as 25. **The image is then represented in a 3D feature space by these 3 histograms.** The location of each training image in this feature space is calculated in the beginning.

**b.** At the time of classification of the test image, k nearest neighbor approach is used. **k has been selected as 5** in the current implementation. So, 5 nearest training images to the present test image are found and the test image is labeled using a majority vote. If, majority of the 5 train images belong a class label, the test image is assigned the same label.

```
%% Compute 3 Channel histogram with 25 bins for each image
for i=1:1888

    im_r=im{i}(:,:,1);
    im_r1=reshape(im_r,[],1);
    hist_train{i,1}=hist(im_r1,25);

    im_g=im{i}(:,:,2);
    im_g1=reshape(im_g,[],1);
    hist_train{i,2}=hist(im_g1,25);

    im_b=im{i}(:,:,3);
    im_b1=reshape(im_b,[],1);
    hist_train{i,3}=hist(im_b1,25);

end
```

**c.** Accuracy obtained by this technique is 44.38%. The low accuracy confirms the initial hypothesis that color histogram is not an efficient and foolproof technique to classify images. The confusion matrix (8 x 8 matrix) is obtained using the ground truth labels of the test images and the classifier's labels. The confusion matrix plot is shown in Fig. 1.

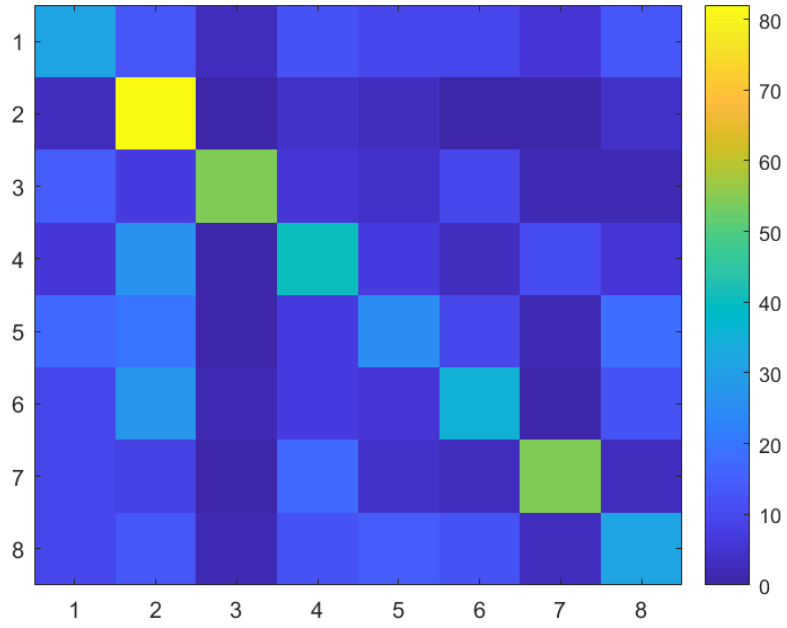


Fig.1 . Confusion Matrix Plot for Color histogram and k-nearest neighbor (kNN) classifier

```
>> C

C =

    32    14     3    12    10     9     6    14
     3    82     0     5     3     1     1     5
    15     7    54     6     5     9     2     2
     6    26     1    40     7     3    11     6
    17    20     1     7    25     9     2    19
     9    28     2     7     6    35     1    12
     9     8     1    17     4     3    55     3
    10    14     2    12    15    12     3    32
```

Fig.2 . Confusion Matrix for Color histogram and k-nearest neighbor (kNN) classifier

## Problem B - Bag of Visual words Model and nearest neighbor classifier

**a.** In this model, we first compute the visual word dictionary. For this computation all the SIFT features from all the training images are combined and a k-Means clustering algorithm is applied. The number of clusters is equal to the vocabulary size desired. In my implementation I have used k = 400. Due to the computational limitations in clustering all the SIFT vectors from all

the images, I have subsampled 30 feature vectors from each of the 1888 training images resulting in a total of 56640 SIFT vectors. **The number of words in the vocabulary, that is the number of clusters has been set at 300.**

b. The algorithm random initializes 300 SIFT vectors as cluster centers and then iterates the k-Means algorithm 20 times, each time assigning the SIFT vectors to its nearest cluster center and re-computing the cluster center. I have used a hard stopping criterion to the k-Means algorithm by setting the total number of iterations at 20. This should be enough for the algorithm to converge. For a better performance, this can be increased at a cost of more computational time.

c. After computing the dictionary, each training image is represented as a word using the 'histc' matlab function. Basically, it's a representation of the number of SIFT features belonging to each of the words in the vocabulary (the SIFT cluster centers). This process is repeated for the test images. Then a kNN classifier is used to assign labels to the test images. I have used k=5 in the current implementation.

d. The accuracy obtained by this technique is 48.13%. This is an improvement over the first approach. The confusion matrix (8 x 8 matrix) is obtained using the ground truth labels of the test images and the classifier's labels. The confusion matrix plot is shown in Fig. 3.

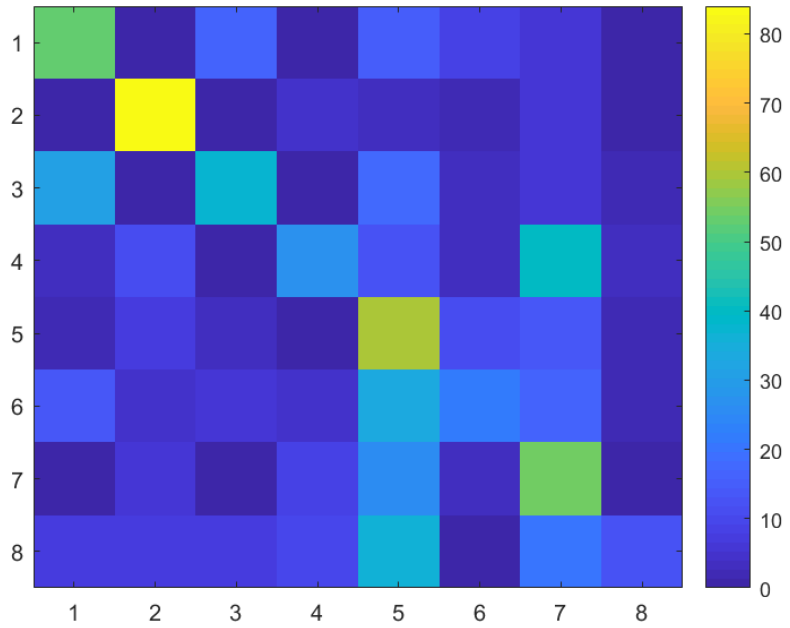


Fig.3 . Confusion Matrix Plot for Bag of Visual words and k-nearest neighbor (kNN) classifier

```
C =

    53     1    16     0    15     9     6     0
     0    84     0     4     3     2     6     1
    31     1    38     1    18     3     6     2
     3    11     0    27    13     3    40     3
     2     7     3     1    60    11    14     2
    14     4     6     4    33    21    16     2
     0     6     0     9    26     3    55     1
     7     7     7    10    36     1    20    12
```

Fig. 4 . Confusion Matrix for Bag of Visual words and k-nearest neighbor (kNN) classifier

## Problem C - Bag of Visual words Model and a discriminative classifier

a. This approach is similar to Problem B. Instead of a nearest neighbor classifier a multiclass SVM classifier is used. After the computation of the visual dictionary and assigning words to all the training and testing images, the SVM is used in a One Vs Rest fashion to classify the test images.

b. As there are a total of 8 classes, 8 SVM models are created. Each image is passed through all the 8 models, and the score from the model of the test image being labeled positive is stored. The image is then assigned the label corresponding to the SVM which assigns it with the highest score.

**c. The time required for training after computing the visual word for all training images is 2.289025 seconds. The time required for testing is 5.509856 seconds.**

d. The accuracy obtained by this technique is 51.00%. This is an improvement over the nearest neighbor approach. The confusion matrix (8 x 8 matrix) is obtained using the ground truth labels of the test images and the classifier's labels. The confusion matrix plot is shown in Fig. 6.

```
C =

    78     2     3     0     9     4     2     2
     0    95     0     1     2     1     1     0
    49     2    20     5     7     8     3     6
     1    24     1    54     1     4     3    12
     6    14     5     3    48     8     5    11
    27    11     1     3    18    27     4     9
     2    12     4     8    28     6    22    18
     8    10     3    10     8     4     3    54
```

Fig. 5 . Confusion Matrix for Bag of Visual words SVM classifier

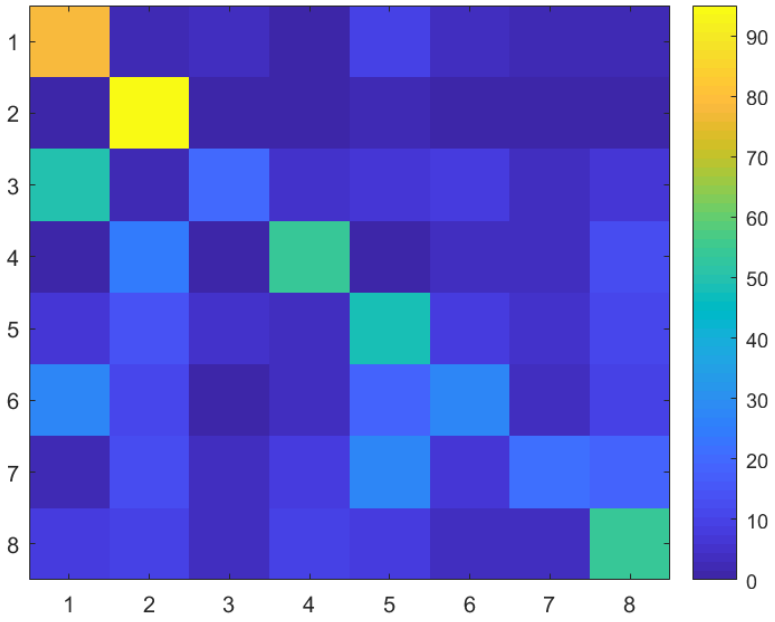


Fig. 6 . Confusion Matrix Plot for Bag of Visual words SVM classifier

## Problem D - Spatial pyramid model and a discriminative classifier

References: Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories, CVPR 2016 and

a. Bag of words discards information about the spatial structure of the image and this information can aid on classification. Spatial pyramid matching divides the image into a small number of cells, and concatenates the histogram of each of these cells to the histogram of the original image, with a suitable weight.

b. **In my implementation I have used L=2, that is a total of 3 layers.** The number of words in the dictionary is set at 200. I have assumed first two rows of train_F and test_F to contain the position of the SIFT vector in the image (all values are < 256 and the images are 256 x 256). At layer 3, the SIFT vectors of each image are divided into 16 parts depending on its position in the image. Then, each cell is considered as a small image and count how often each visual word appears. Finally to represent the entire image, we concatenate all the histograms together after normalization by the total number of features in the image. The weights assigned are ¼, ½ and ¼ for Layer 0, Layer 1 and Layer 2 respectively.

c. This method leads to each image being represented by a 4200 x 1 vector (200 * (1+4+16)).

d. Afterwards, a multiclass SVM model is used to classify the images. As there are a total of 8 classes, 8 SVM models are created.

e. The accuracy obtained by this technique is 59.08%. The confusion matrix (8 x 8 matrix) is obtained using the ground truth labels of the test images and the classifier's labels. The confusion matrix plot is shown in Fig. 8.

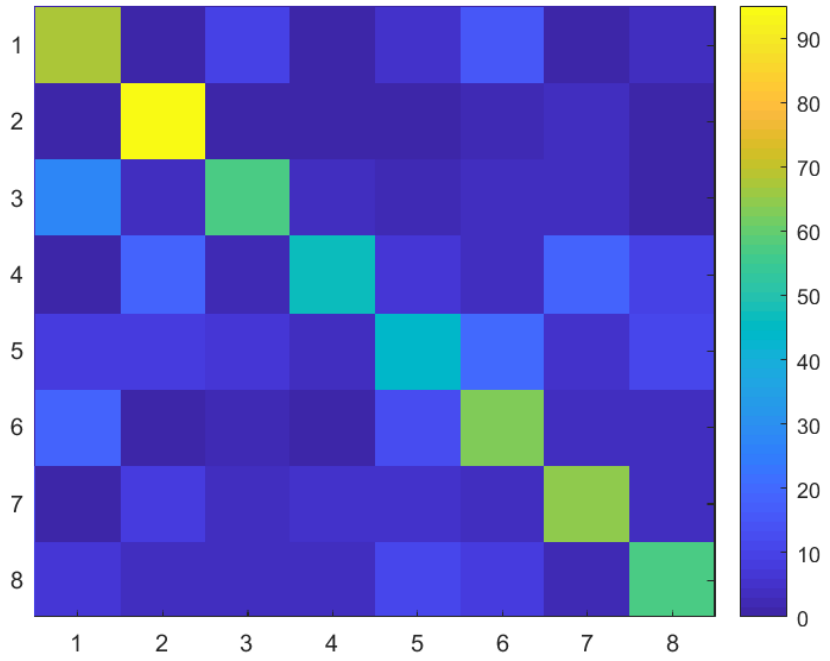


Fig. 7 . Confusion Matrix Plot for Spatial Pyramid Model and Discriminative Classifier

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 67 | 1 | 6 | 1 | 4 | 16 | 1 | 1 |
| 0 | 87 | 0 | 0 | 3 | 1 | 4 | 2 |
| 22 | 0 | 56 | 4 | 3 | 2 | 5 | 3 |
| 1 | 16 | 1 | 42 | 6 | 6 | 13 | 14 |
| 7 | 5 | 6 | 5 | 35 | 23 | 7 | 11 |
| 14 | 0 | 6 | 1 | 14 | 60 | 3 | 1 |
| 1 | 6 | 1 | 9 | 3 | 5 | 60 | 5 |
| 4 | 4 | 2 | 8 | 3 | 8 | 2 | 60 |

Fig. 8 . Confusion Matrix for Spatial Pyramid Model and Discriminative Classifier

## Graduate Points

### 1. Fisher Vector Encoding

Reference : The devil is in the details: an evaluation of recent feature encoding methods, BMVC 2011

a. VLFeat Bianry package is installed to use Fisher encoding methods. After downloading the toolbox is set up.

```
run('vlfeat-0.9.20/toolbox/vl_setup');
```

b. The words corresponding to each of the training and test images are formed according to the Fisher Encoding scheme.

```
%% Computing word for each Train Image
for i = 1:length(train_gs)
encoding = vl_fisher(double(train_D{i}), clusterMeans,
clusterCovariances, clusterPriors);
trainData(i, :) = encoding;
end
```

c. Afterwards, a multiclass SVM model is used to classify the images. As there are a total of 8 classes, 8 SVM models are created.

f. The accuracy obtained by this technique is 63.375%. The confusion matrix (8 x 8 matrix) is obtained using the ground truth labels of the test images and the classifier's labels. The confusion matrix plot is shown in Fig. 10 and the matrix is shown in Fig 9.

```
C =

    56     0     7     1    15    17     2     2
     0    86     0     1     7     3     0     3
    17     1    48     2     9    11     7     5
     0     2     0    75     1     4     5    13
     4     7     1     0    65    16     3     4
    20     6     3     0    16    51     3     1
     2     2     0    15     9     7    60     5
     1     0     0    13     6     4    10    66
```

Fig. 9 . Confusion Matrix for Fisher Encoding and Discriminative Classifier Model

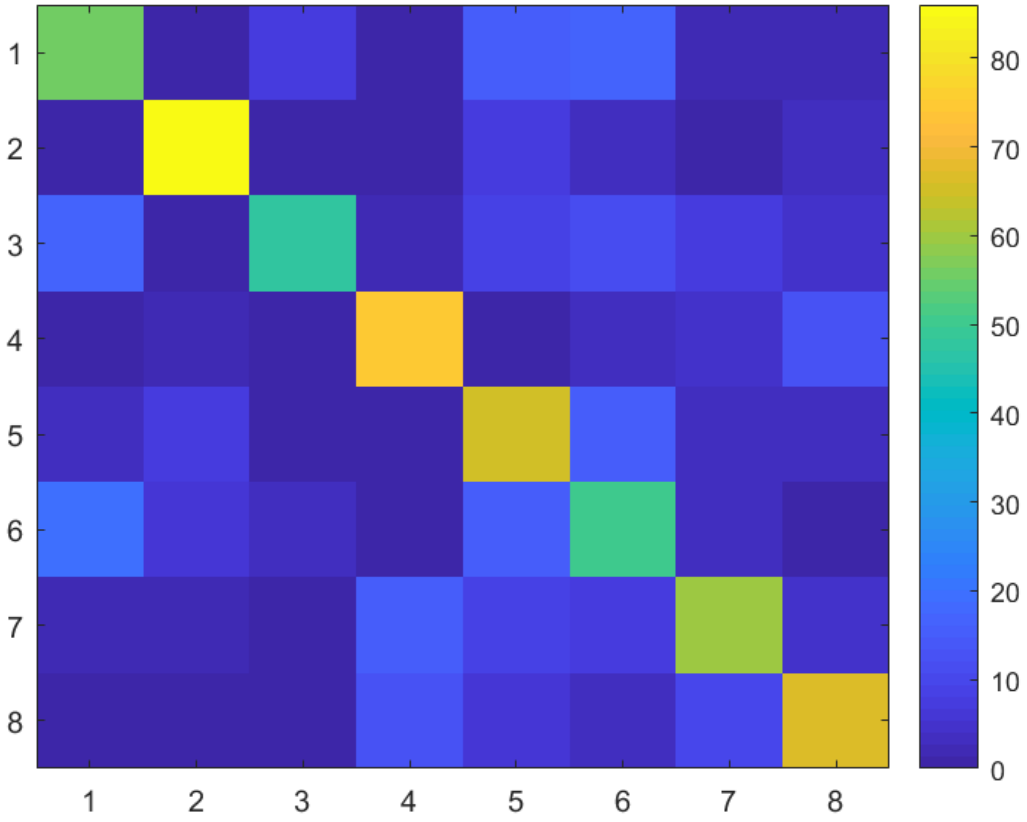


Fig. 9 . Confusion Matrix Plot for Fisher Encoding Model and Discriminative Classifier

## 2. Using different number of Visual Words in Bag of Words Model

I tried out different vocabulary size in the bag of words model by changing the parameter K in the 'runThisq2.m' code. These are the observed categorization accuracies:

- K= 1000. Accuracy is 49.06%
- K= 400. Accuracy is 48.13%
- K= 300. Accuracy is 46.75%
- K = 600. Accuracy is 47.13%
- K = 50. Accuracy is 48.75%
- K =100. Accuracy is 48.50%

So, we observe that there are marginal differences in accuracies by using different dictionary size.