Key points for code submission:

1. batchToVisualWords() has been modified to save files for random and Harris wordMaps in different directories.
2. The data_random folder contains the .mat files for random wordMaps while data folder contains the .mat files for Harris wordMaps.
3. buildRecognitionSystem() will pick up the .mat files from these directories.

Q1.1 What properties do each of the filter functions pick up? You should group the filters into broad categories (i.e all Gaussians, derivatives etc).

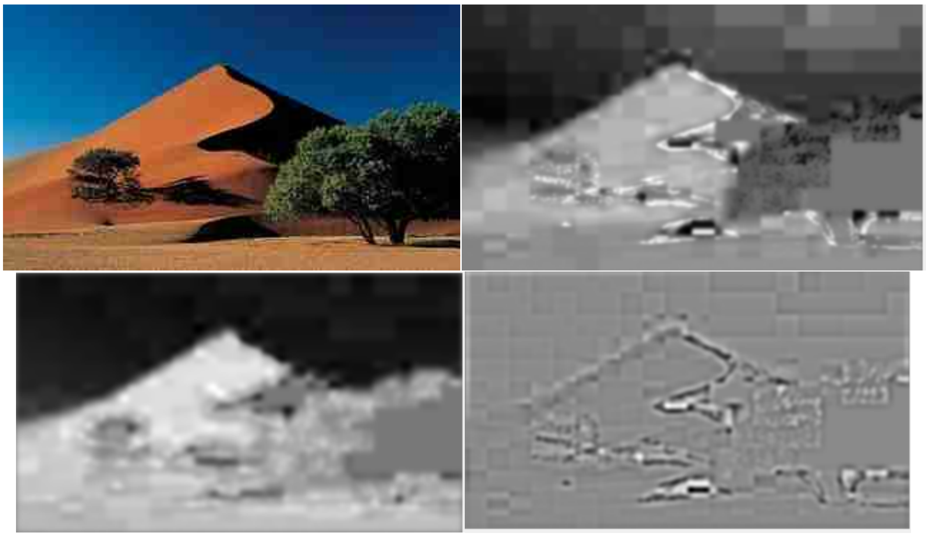
Ans. The filters belong to mainly the following four categories:

(a) Gaussian filter
(b) Derivative-x filter
(c) Derivative-y filter
(d) Laplacian of Gaussian filter

The Gaussian filters help reduce the noise in the image and cause blurring, used as a pre-smoothing filter.
The first order derivative filters are used for edge detection in the image. But the edges may be difficult to localize precisely.
The derivative-x filter enhances edges in the x-direction while the derivative-y filter enhances edges in the y-direction.
The Laplacian of Gaussian filters are second order derivative filters used for highlighting rapid change in intensity values and is the zero crossing edge detector. It is passed through the Gaussian filter for smoothing the image as the derivative filters amplify noise.

Q1.2 Show an image from the dataset and 3 of its filter responses. Explain any artifacts you may notice. Also briefly describe the CIE Lab colorspace, and why we would like to use it.

Ans.



Sample desert image with 3 filter responses.

The effect of the filters on these images is enhancement of edges and lowpass filtering leading to reduction of high frequency elements. For the responses shown above, we can see a blurred image which could be a Gaussian filter response and other two have sharp edges, being derivative filter responses. The edges are being detected through the Laplacian and derivative filters.

The CIE Lab colorspace consists of three channels – L for luminance, two color channels a and b. This is a three-dimensional color space where each color difference corresponds to distances when measured colorimetically. It is mainly used for sharpening images and perceiving difference in colors better that aids in image processing.

Q1.3 Show the results of your corner detector on 3 different images.

Ans.

Image 1

Image 2

Image 3

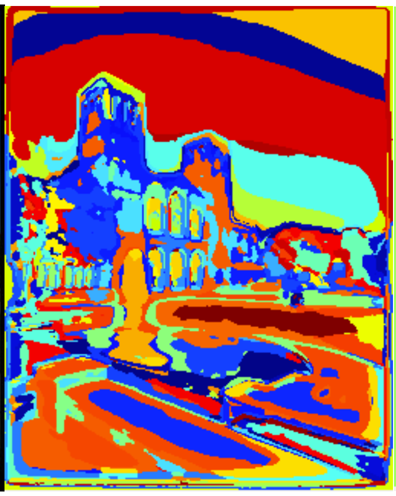
Q1.4 Compute Dictionary of Visual Words

Ans. Please refer to the code.

Q2.1 Show the wordmaps for 3 different images from two different classes (6 images total). Do this for each of the two dictionary types (random and Harris). Are the visual words capturing semantic meanings? Which dictionary seems to be better in your opinion? Why?

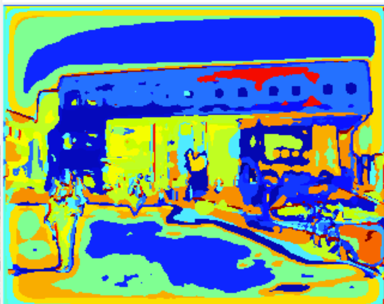
Ans.

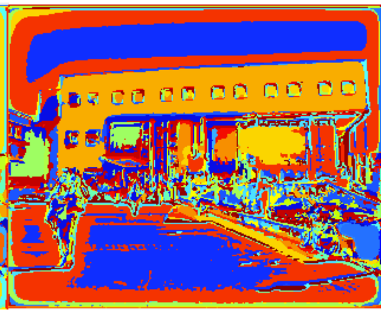

Campus image 1

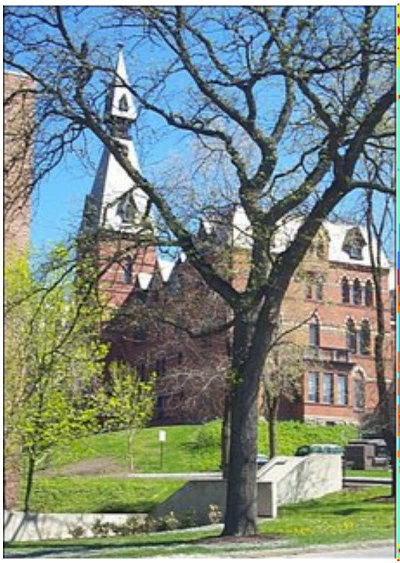

Random

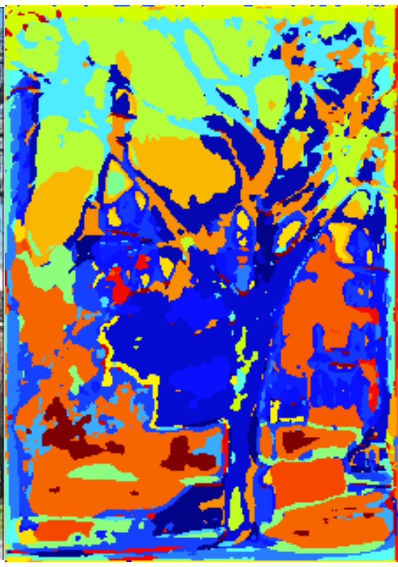

Harris

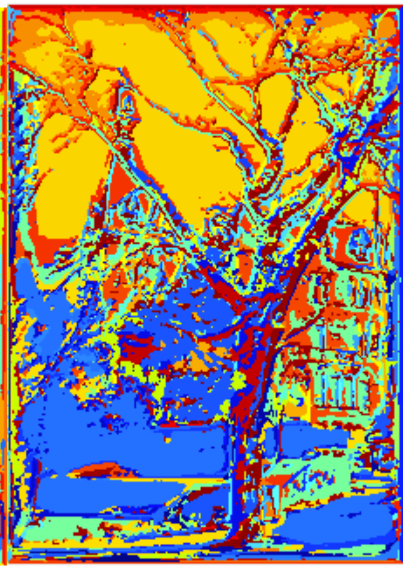

Campus image 2


Random

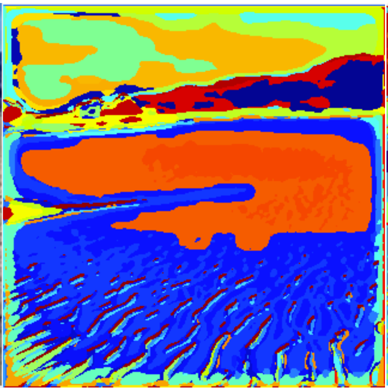

Harris


Campus image 3

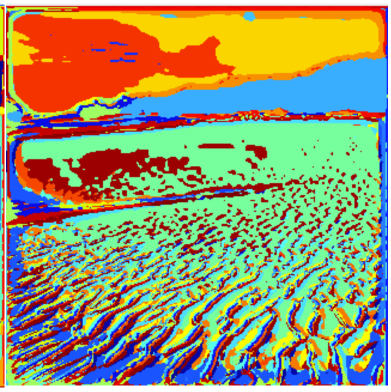

Random


Harris

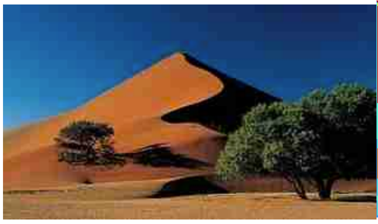
Desert image 1



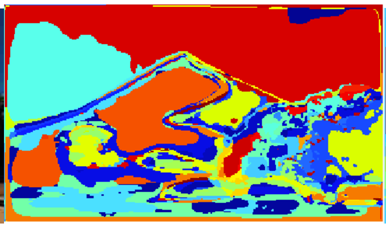
Random



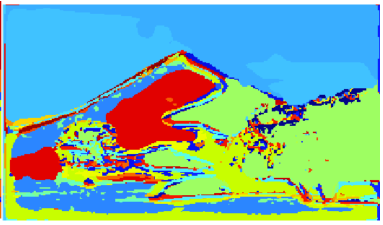
Harris



Desert image 2

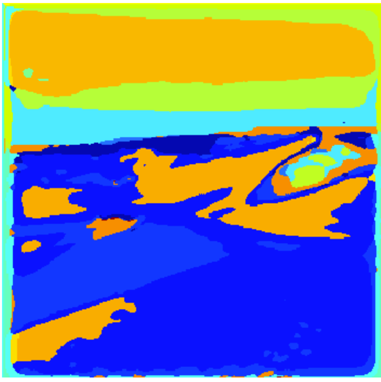


Random

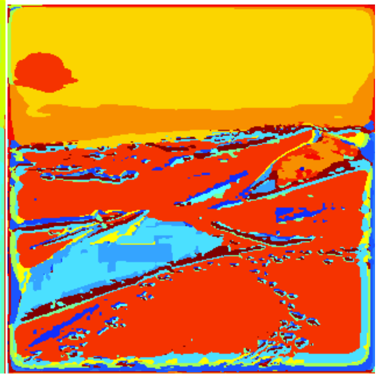


Harris

Desert image 3



Random



Harris

The visual words do seem to have a semantic meaning, for Harris detector especially as it captures more details. The Harris dictionary seems to be better than the random dictionary as it preserves the corners of the image in detail which could be useful in image classification.

Q2.2 Get Image Features

Ans. Please refer to code.

Q2.3 Build Recognition System - Nearest Neighbours

Ans. Please refer to code.

Q3.1 Image Feature Distance

Ans. Please refer to code

Q3.2 Comment on whole system:

- Include the output of evaluateRecognitionSystem.m (4 confusion matrices and accuracies).
- How do the performances of the two dictionaries compare? Is this surprising?
- How about the two distance metrics? Which performed better? Why do you think this is?

Ans.

Random, Euclidean

Accuracy: 0.4750

C:

12	1	3	1	0	0	2	1
3	13	1	0	1	2	0	0
5	4	9	0	1	1	0	0
1	2	3	9	1	1	2	1
3	1	1	3	9	0	3	0
1	1	2	2	1	5	4	4
3	1	0	6	3	1	4	2
3	1	0	1	0	0	0	15

Random, Chi2

Accuracy: 0.5312

C:

13	1	3	0	0	0	1	2
3	12	3	0	1	1	0	0
4	3	13	0	0	0	0	0
1	2	3	6	2	2	2	2
1	1	1	2	11	1	3	0
2	2	2	1	1	7	2	3
5	1	1	3	3	0	7	0
2	0	0	1	0	0	1	16

Harris, Euclidean

Accuracy: 0.4188

C:

8	2	5	0	0	1	2	2
5	11	2	1	1	0	0	0
5	3	8	2	1	1	0	0
1	3	0	10	0	1	4	1
0	2	3	3	5	0	6	1
0	2	1	4	1	4	4	4
3	1	3	5	0	2	4	2
2	0	0	1	0	0	0	17

Harris, Chi2

Accuracy: 0.5188

C:

15	1	1	0	0	0	0	3
3	11	3	0	1	1	0	1
3	4	10	0	3	0	0	0
2	2	1	6	0	1	7	1
0	2	1	1	13	0	2	1
0	3	2	3	1	5	3	3
1	0	4	3	0	1	9	2
4	0	0	2	0	0	0	14

Random dictionary performs slightly better than Harris dictionary for both Euclidean and Chi2 metrics. This is surprising because Harris is expected to perform well as it detects corners. This could be a case where random has chosen better values by chance and hence computed a slightly better accuracy than Harris dictionary.

Chi2 metrics perform better than Euclidean, as observed for this dataset. This is primarily because Chi2 is a weighted Euclidean. It accounts higher weight for species whose total abundance in the data is low and lowers priority of abundant species. It thus tends to exaggerate the distinctiveness of samples containing several rare species which is better for our classifier.

QX.1 Evaluate Recognition System - Support Vector Machine

Ans.

Using RBF kernel,

accuracyHarris = 0.4750
C:

14	4	1	0	0	0	0	1
4	14	0	0	1	0	1	0
2	7	11	0	0	0	0	0
3	2	1	7	0	0	7	0
1	5	1	0	7	0	6	0
4	3	3	5	0	0	4	1
5	0	1	1	3	0	10	0
4	1	0	1	0	0	1	13

accuracyRandom = 0.5375
C:

15	3	1	0	0	0	0	1
5	13	0	0	1	0	1	0
2	7	11	0	0	0	0	0
3	2	0	8	0	1	6	0
0	3	2	0	11	0	4	0
4	3	2	2	1	4	3	1
4	0	0	1	5	0	9	1
3	1	0	0	0	0	1	15

Using Linear kernel,

accuracyHarris = 0.4437
C:

13	4	2	0	0	0	0	1
4	14	1	0	0	0	1	0
2	5	11	0	2	0	0	0
3	2	1	6	0	0	8	0
0	5	2	0	6	0	7	0
4	3	3	5	0	0	4	1
6	0	1	1	4	0	8	0
5	1	0	1	0	0	0	13

accuracyRandom = 0.4562
C:

14	4	1	0	0	0	0	1
4	13	0	0	2	0	1	0
2	7	11	0	0	0	0	0
3	2	0	6	0	0	9	0
0	5	3	1	5	0	6	0
5	3	2	1	2	1	5	1
6	0	0	1	4	0	8	1
3	1	0	0	0	0	1	15

The SVM classifier was tested on the given dataset using two kernels, “RBF” and “linear”. The SVM classification based on RBF kernel in general performed better than the nearest neighbor classifier. It was more accurate than Harris Euclidean, but was less accurate than Harris Chi2. For Random, it was more accurate than both Euclidean and Chi2. The performance of various classification methods depends greatly on the general characteristics of the data to be classified.

The RBF kernel worked better than the linear kernel in this case. Linear kernel works best when data is linearly separable. RBF is a non-linear classifier and works well when the data set and feature set is small enough.

QX.2 Inverse Document Frequency

Ans.

accuracyEuclideanHarris = 0.3750

C:

10	0	6	1	0	0	1	2
8	5	6	0	1	0	0	0
4	5	9	1	1	0	0	0
1	1	2	7	1	3	5	0
1	2	3	1	4	4	5	0
2	3	2	3	2	6	0	2
1	2	2	1	7	2	4	1
3	0	0	2	0	0	0	15

accuracyChi2Harris = 0.4813

C:

9	1	4	2	0	0	1	3
6	7	3	0	1	1	1	1
2	7	10	0	1	0	0	0
2	1	1	7	1	4	4	0
0	0	4	2	10	0	3	1
1	3	1	4	1	5	2	3
1	1	1	0	3	0	12	2
1	0	0	0	0	1	1	17

accuracyEuclideanRandom = 0.4250

C:

9	4	4	1	0	0	0	2
7	9	2	0	1	1	0	0
6	5	7	0	1	1	0	0
1	3	2	5	2	2	4	1
1	1	1	3	13	0	1	0
2	2	0	5	2	5	1	3
3	1	1	3	3	2	5	2
1	0	1	2	0	1	0	15

accuracyChi2Random = 0.4938

C:

10	1	5	1	0	0	0	3
4	11	3	0	1	1	0	0
2	5	11	0	1	0	1	0
0	2	2	8	2	2	3	1
2	1	2	2	9	1	3	0

0	2	2	3	1	7	2	3
1	1	0	4	3	1	8	2
0	0	0	3	0	1	1	15

The Inverse Document Frequency should perform better as the frequent words become insignificant for image classification and more weightage is given to unique words. This helps in correctly classifying the test images. In this case, however, there is decrease in accuracy of about 5% for both Random and Harris dictionaries with IDF.