16720 B

HW2 - BoVW

AndrewID – prakharb

Q1.1.1 (5 Points WriteUp)

What visual properties do each of the filter functions (See Figure below) pick up? You should group the filters into categories by its purpose/functionality. Also, why do we need multiple scales of filter responses?

(1) Gaussian – This is a low pass filter which we use for reducing the noise i.e. the high frequency components of the image and it is also used for image blurring.

(2) Laplacian of Gaussian – This is the third member of the family of 2D gaussian filters. We use this filter for detecting the edges of an image as it calculates the intensity changes in an image.

(3) derivative of Gaussian in the xx direction - Derivative of Gaussian in the xx direction is the second member of the family of 2D gaussian filters. Helps detect the intensity changes in the x direction.

(4) derivative of Gaussian in the yy direction - Derivative of Gaussian in the yy direction is the second member of the family of 2D gaussian filters. Helps detect the intensity changes in the y direction.

We need multiple scales for the filter responses as when we apply the same filter with different scales on an image then that is similar to applying the same filter with different scales on the sub sampled image.

Answer 1. Harris Corners

Image 1

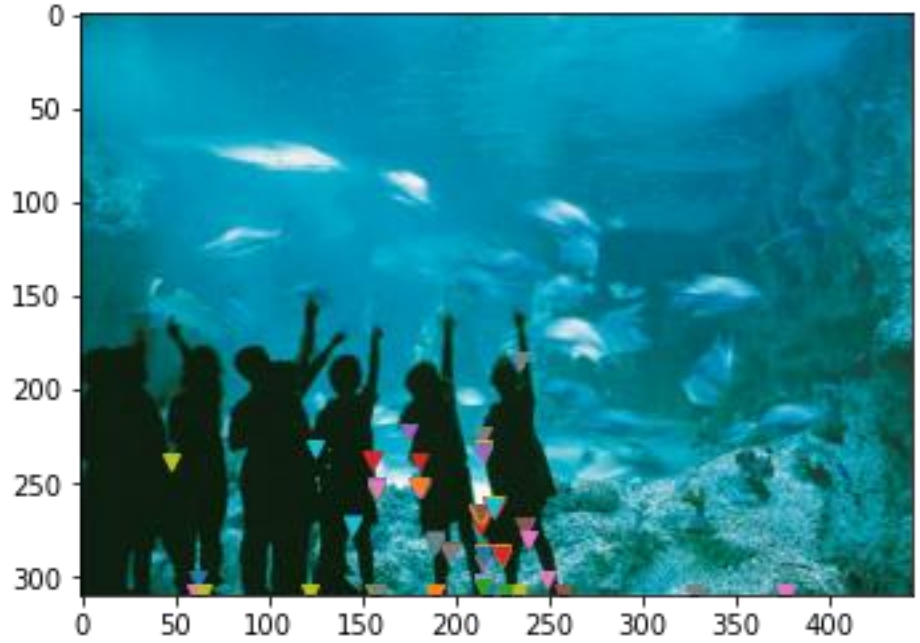


Image 2

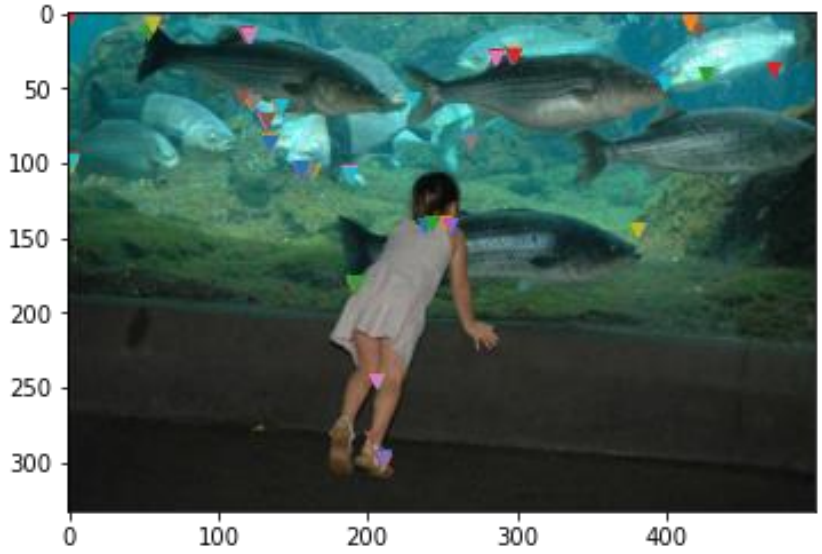


Image 3

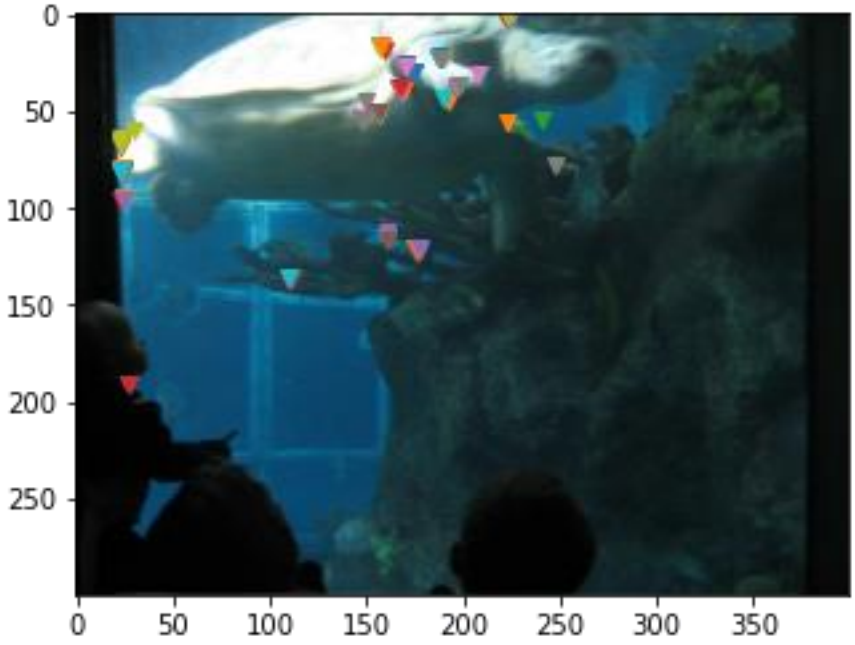


Image 4

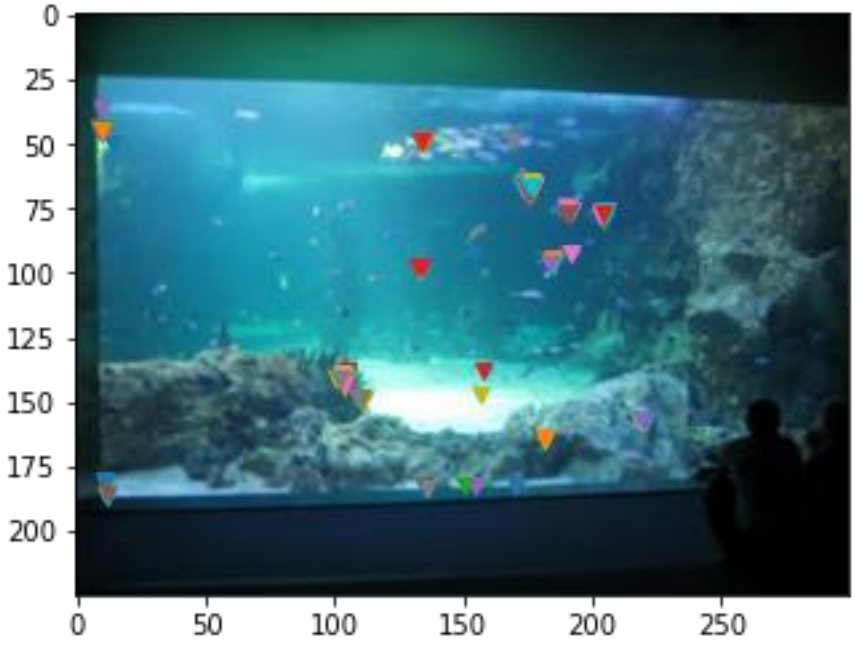
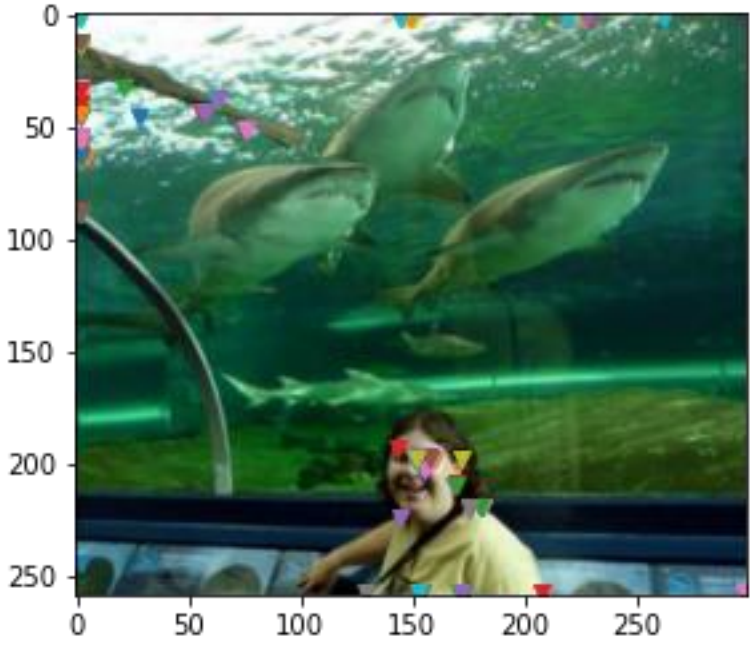


Image 5



1.3.1 Word Maps

Image1

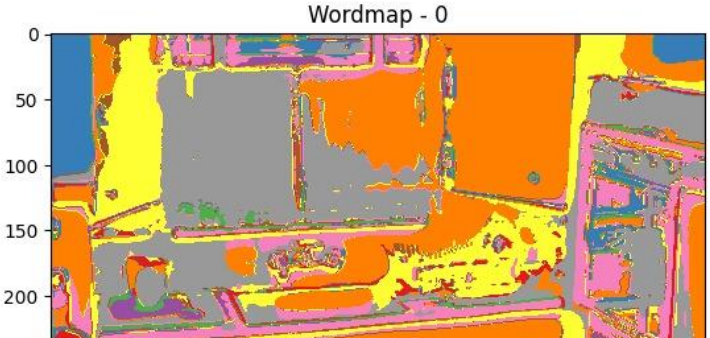


Image2

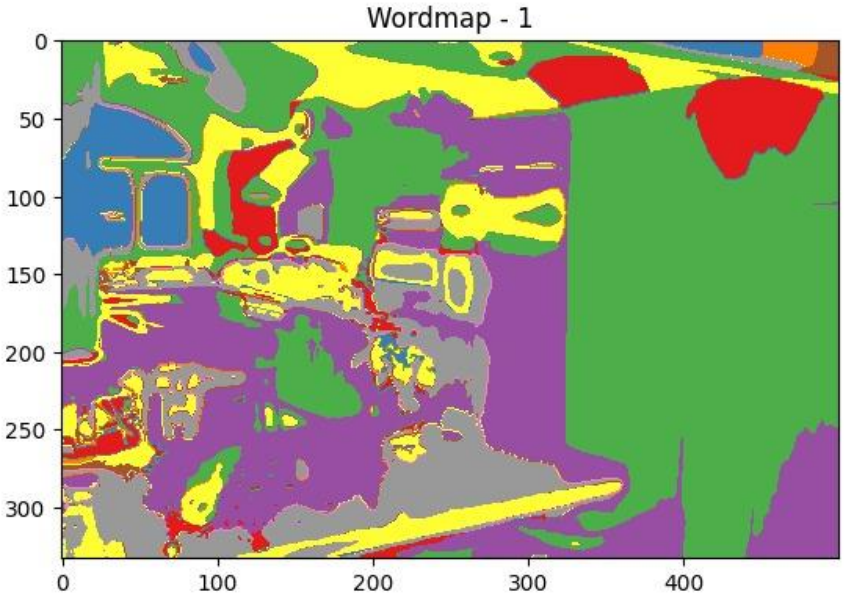
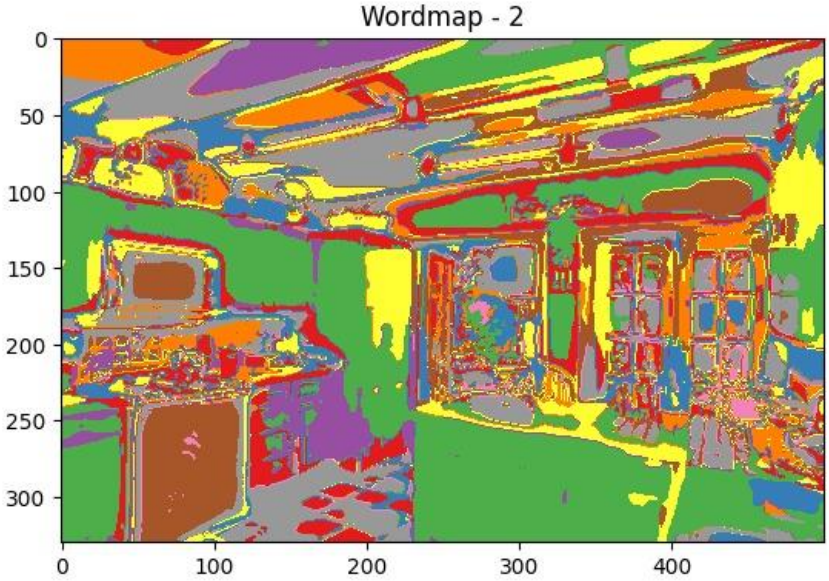


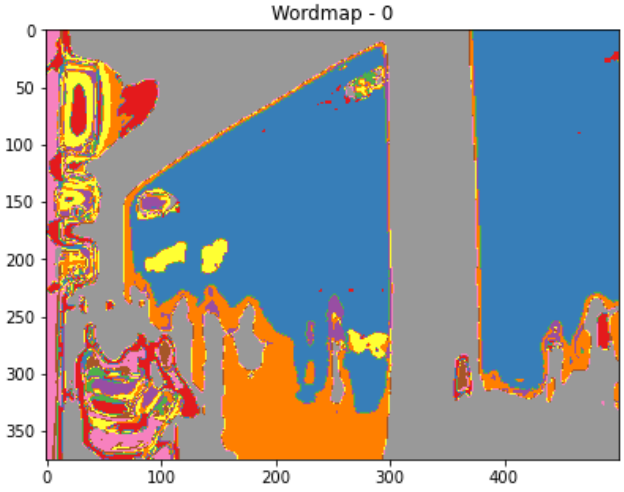
Image3



The word boundaries do make sense. When we compare the outputs with the original images, it is clearly visible that different objects have been differentiated in the image

Answer 2.1

Image1



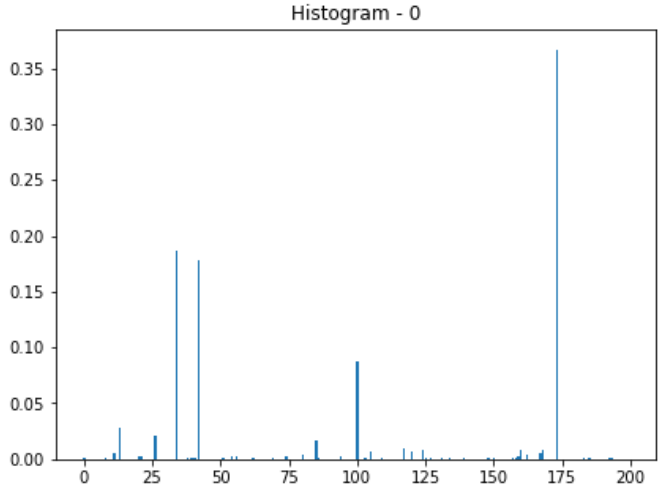
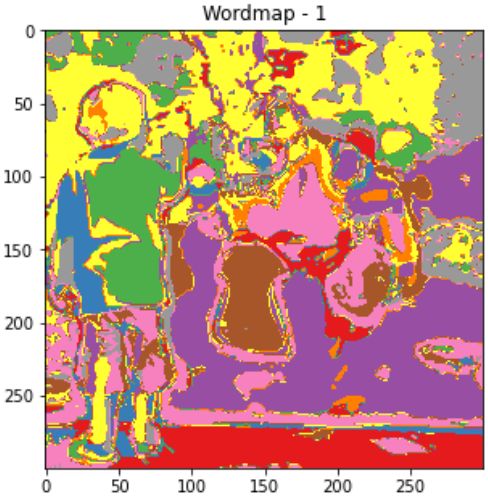


Image2


Wordmap - 1
0
50
100
150
200
250
0
50
100
150
200
250

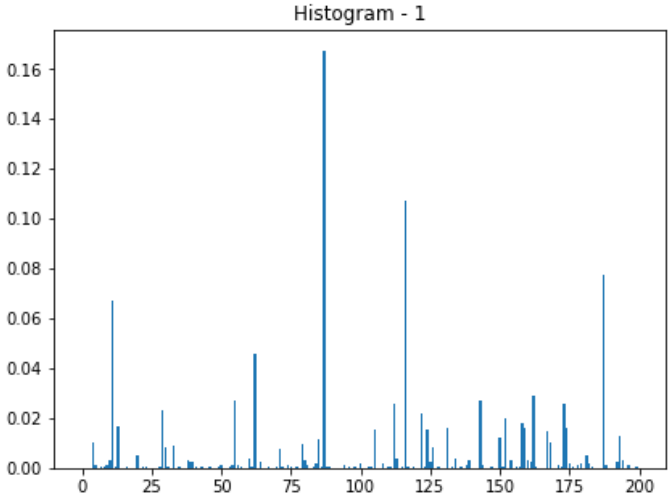
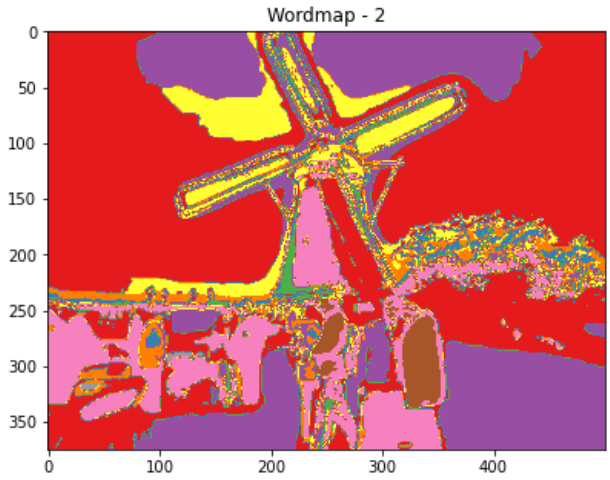

Histogram - 1
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
0
25
50
75
100
125
150
175
200

Image3


Wordmap - 2
0
50
100
150
200
250
300
350
0
100
200
300
400

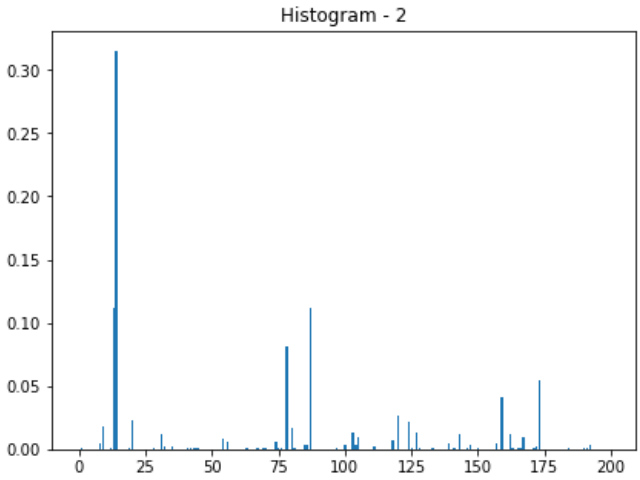
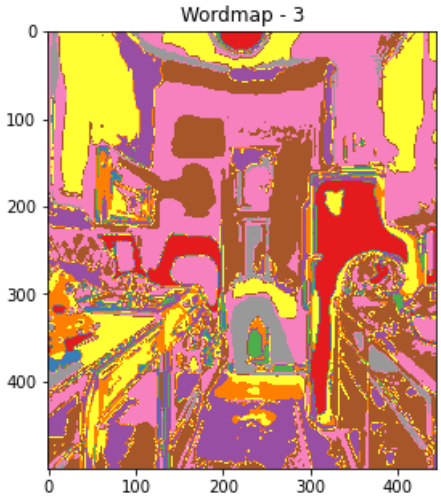

Histogram - 2
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0
25
50
75
100
125
150
175
200

Image4


Wordmap - 3
0
100
200
300
400
0
100
200
300
400

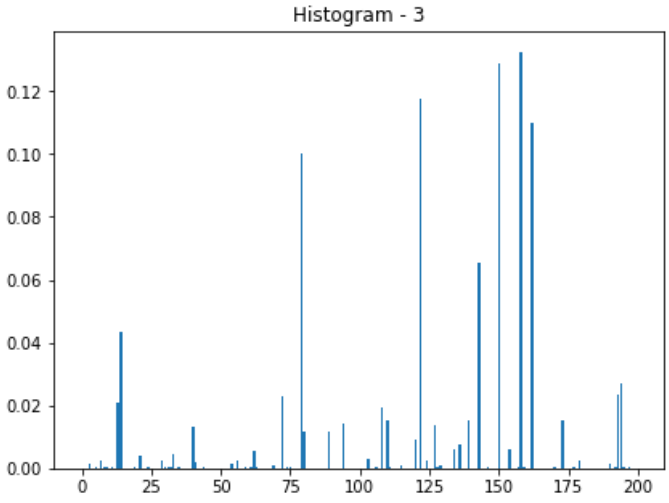
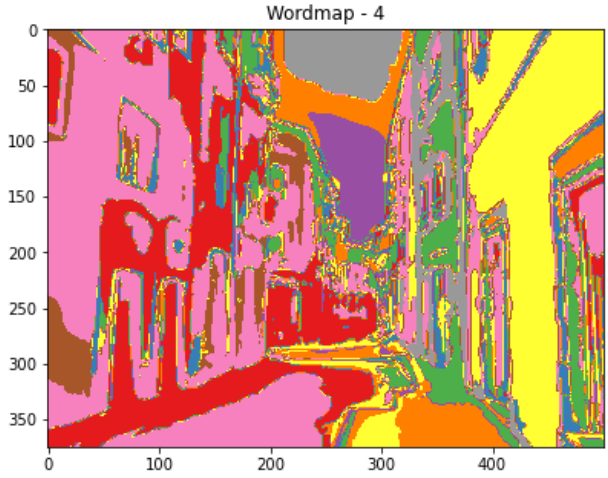
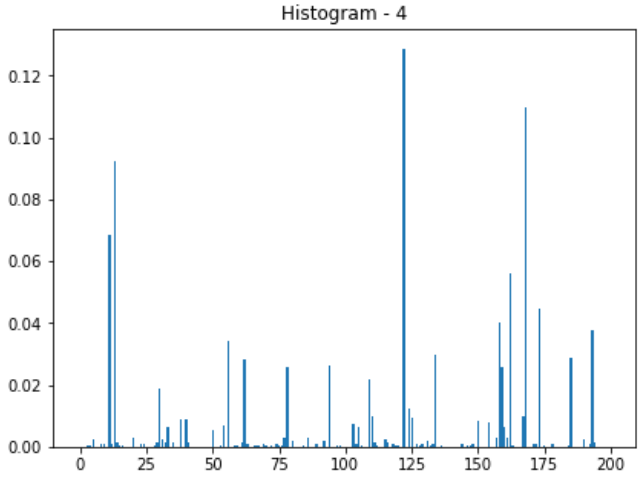

Histogram - 3
0.12
0.10
0.08
0.06
0.04
0.02
0.00
0
25
50
75
100
125
150
175
200

Image5


GIRONA


Wordmap - 4
0
50
100
150
200
250
300
350
0
100
200
300
400


Histogram - 4
0.12
0.10
0.08
0.06
0.04
0.02
0.00
0
25
50
75
100
125
150
175
200

Answer 3.1.1

```
Accuracy: 0.6
accuracy = 0.6
conf_mat = array([[14.,  0.,  0.,  0.,  0.,  0.,  0.,  0.],
       [ 0., 14.,  0.,  0.,  0.,  1.,  2.,  1.],
       [ 0.,  0., 15.,  3.,  4.,  1.,  0.,  2.],
       [ 1.,  3.,  2., 10.,  0.,  3.,  1.,  6.],
       [ 2.,  0.,  0.,  0., 10.,  1.,  0.,  0.],
       [ 1.,  1.,  0.,  1.,  7., 12.,  2.,  0.],
       [ 0.,  3.,  1.,  1.,  3.,  2., 11.,  0.],
       [ 0.,  1.,  1.,  4.,  2.,  0.,  1., 10.]])
```

Answer 3.1.2

0	1	2	3	4	5	6	7
aquarium	park	desert	highway	kitchen	laundromat	waterfall	windmill

In the array above, when we look at the row 1, 14 images are classified as aquarium and no image of aquarium is mis-classified.

For park images 14 of them are correctly classified while 1 is classified as laundromat, 2 are classified as waterfall, 1 is classified as windmill.

For windmill images 10 are classified as correct, 1 image is classified as park, desert. 4 as highway, 2 as kitchen, 1 as waterfall. In this case the classification is done incorrectly into other categories.

Similar is seen for the images of the category of waterfall and laundromat and highway

Reasons –

1. View-Point Variation – While looking at the image set of the windmills it's clear that there are too many variations in the viewpoint, where the windmills are oriented multiple directions. With this kind of orientation and view point variation mis classification is bound to happen.

2. Illumination – Image features may appear to be lost or different with different illumination. This might have caused the misclassification as the brightness levels of the image set may be quite different and inappropriate.

3. Background Clutter – Too much of background noise due to the background clutter may be another reason for the misclassification.

Answer 4.2.1

```
Accuracy: 0.975
Confusion Matrix [[14.  0.  0.  0.  0.  0.  0.  0.]
 [ 0. 17.  0.  0.  0.  0.  0.  1.]
 [ 0.  0. 24.  0.  0.  0.  0.  1.]
 [ 0.  0.  0. 26.  0.  0.  0.  0.]
 [ 0.  0.  0.  0. 12.  1.  0.  0.]
 [ 0.  0.  0.  0.  1. 23.  0.  0.]
 [ 0.  0.  0.  0.  0.  0. 21.  0.]
 [ 0.  0.  0.  0.  0.  0.  0. 19.]]
```

Deep learning algorithms are better than the classical Bag of Visual Words as BoVW ignores the spatial relationships among the patches. As the neural networks in deep learning are trained rather than being programmed, we get greater accuracy as compared to the classical CV techniques.