Name: Srinivas Ravi
UB ID: 50244669
Course: CSE-573 : Introduction to Computer Vision and Image Processing

HW1 – Spatial Pyramid Matching for Scene Classification (using Bag-of-Words approach)

Q1.0 (5 points) What properties do each of the filter functions (see Figure 3) pick up? You should group the filters into broad categories (i.e., all the Gaussians). Answer in your write-up.

Filters:

Gaussian Filter:
        The Gaussian Filters picks up general structure of the image. It blurs the noises while still keeping the edges sharper than a regular averaging filter.

Laplacian of Gaussian (LoG) Filter:
        The Laplacian of Gaussian filters pick up edges (areas of rapid change) in images. It is a derivative filter.

Gaussian Cross correlation filters w.r.t. x
        This captures all the vertical lines in an image. To achieve the same [-1 0 1] is used.

Gaussian Cross correlation filters w.r.t. y
        This captures all the horizontal lines. To achieve the same [-1 0 1]' is used.

The properties highlighted by each of the filters is shown in Figure 1.

Q1.1 (10 points) Apply all 20 filters on a sample image, and visualize as an image collage. Submit the collage of 20 images in the write-up.

Figure 1 - Collage of application of the 4 filters (rows) and their respective scales (columns)

Q1.3 (10 points) Visualize three wordmaps of three images from any one of the category and submit in the write-up along with their original RGB image. Also, provide your comments about the visualization.
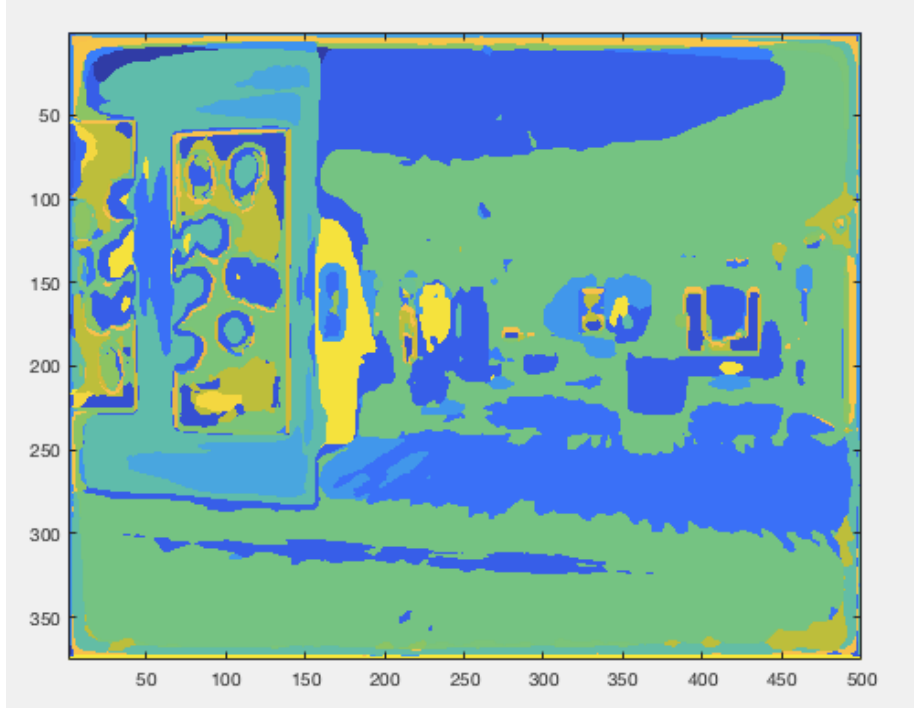
1. Art Gallery:

Figure 2 - Image(top) and the corresponding WordMap(bottom)
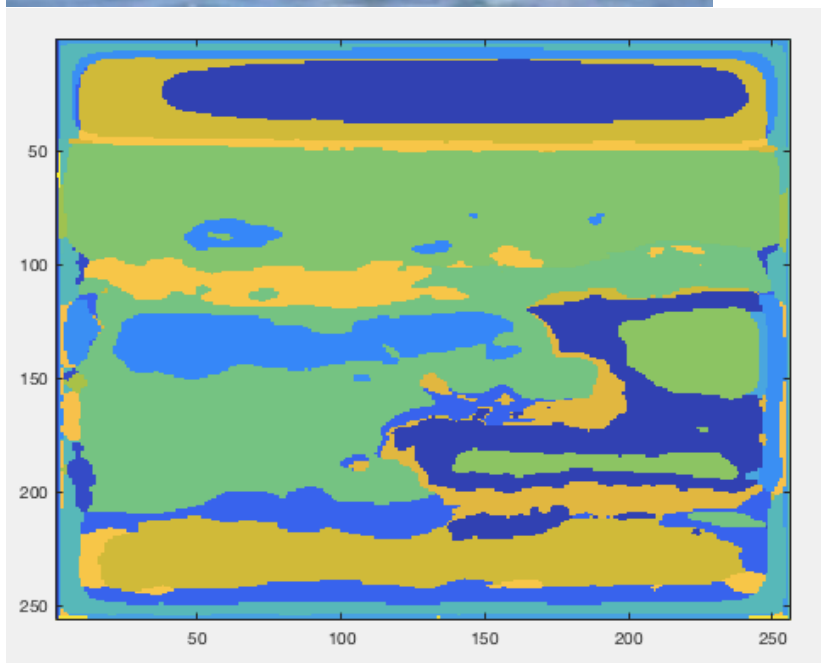
2. Ocean:

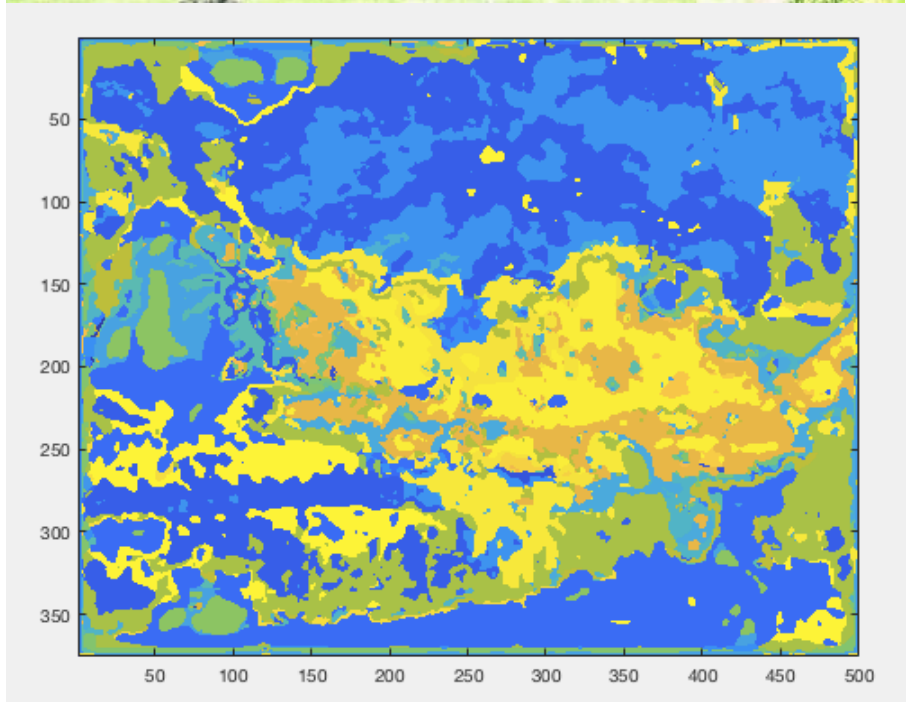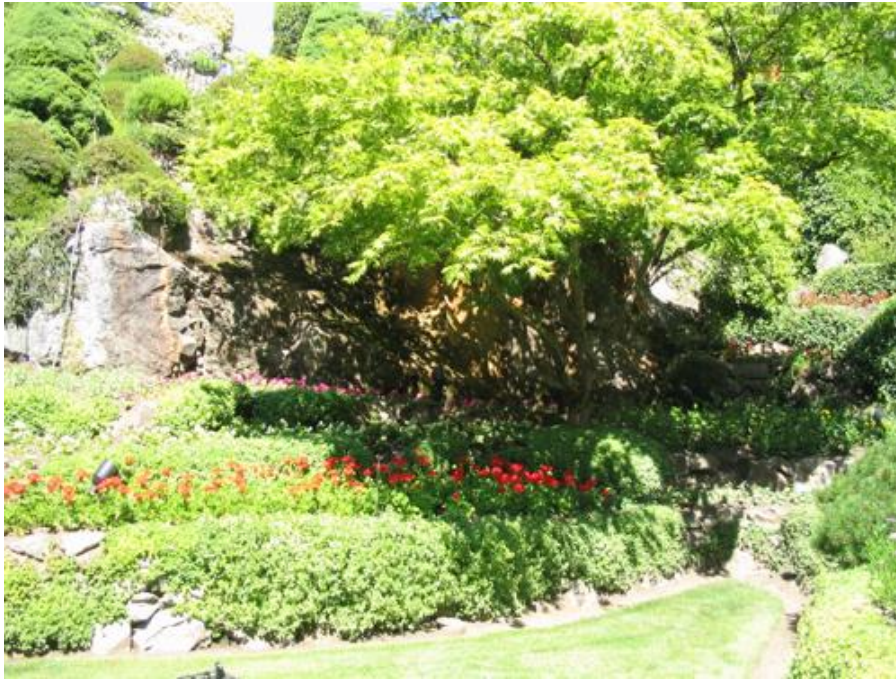*Figure 3 - Image(top) and the corresponding WordMap(bottom)*

3. Garden:

*Figure 4 - Image(top) and the corresponding WordMap(bottom)*

Comment:
The visualization of the wordmap shows that all the pixels in the image are now represented by a word defined in the dictionary. This is going to be used further for scene classification.

Q2.4 Guess image on some of the test samples:


```
>> script1
[Loading..]
[Getting Visual Words..]
[My Guess]:art_gallery.
```

image to classify
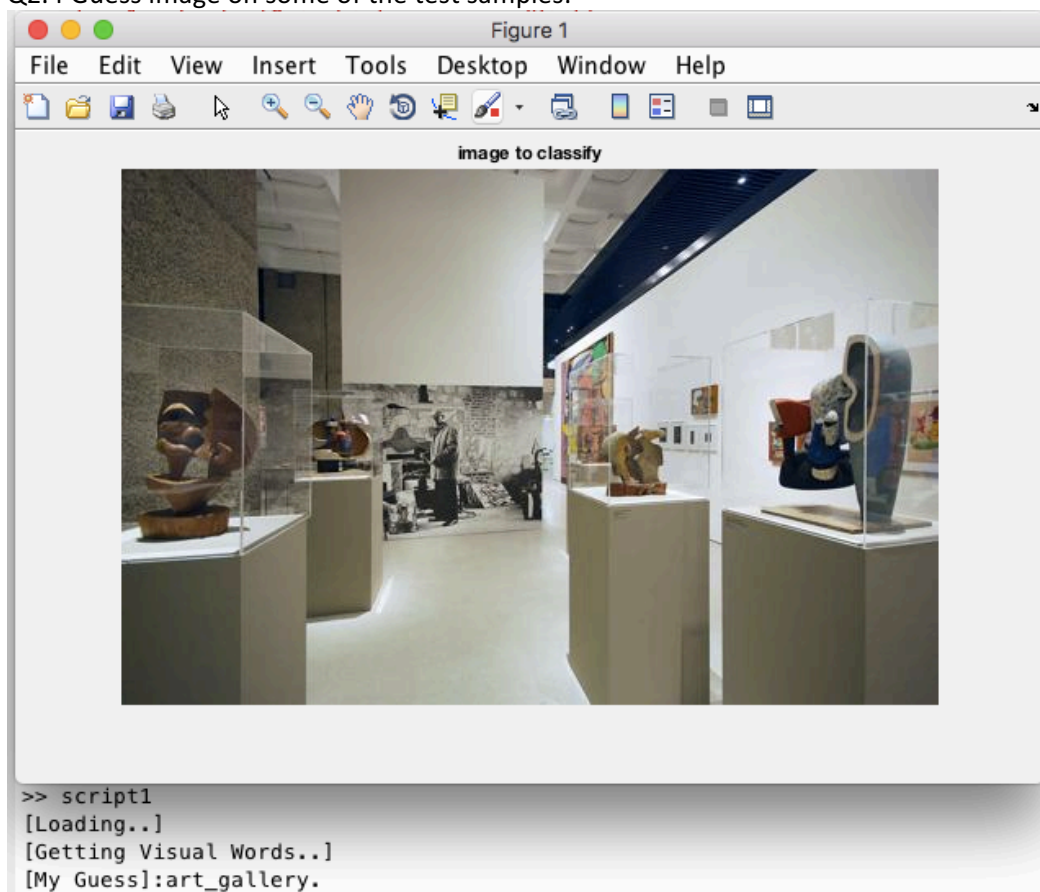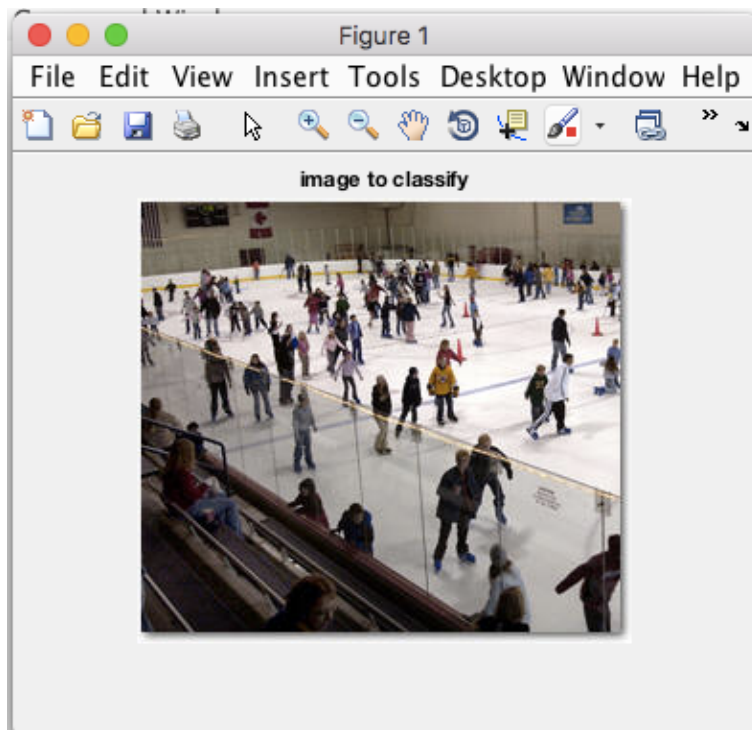
```
>> script1
[Loading..]
[Getting Visual Words..]
[My Guess]:ice_skating.
```


image to classify
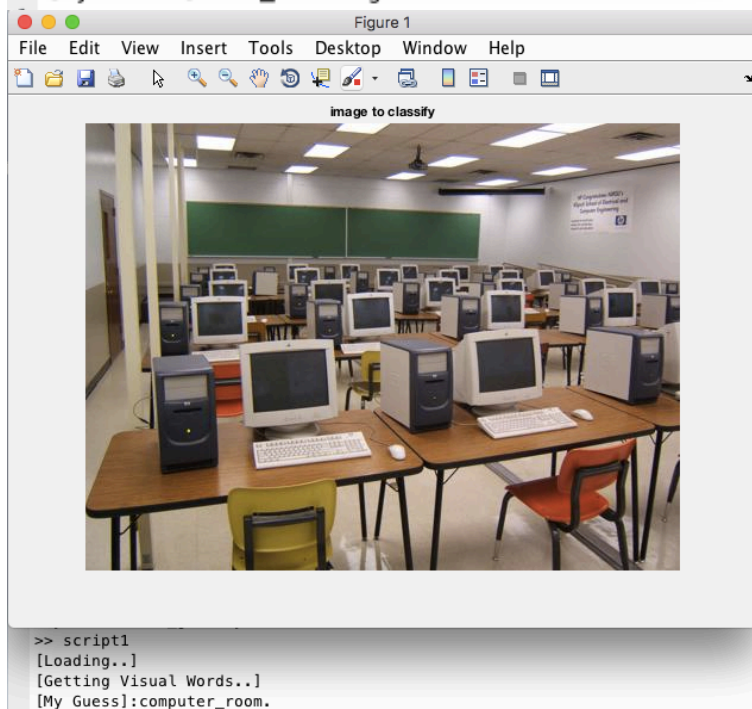
```
>> script1
[Loading..]
[Getting Visual Words..]
[My Guess]:computer_room.
```
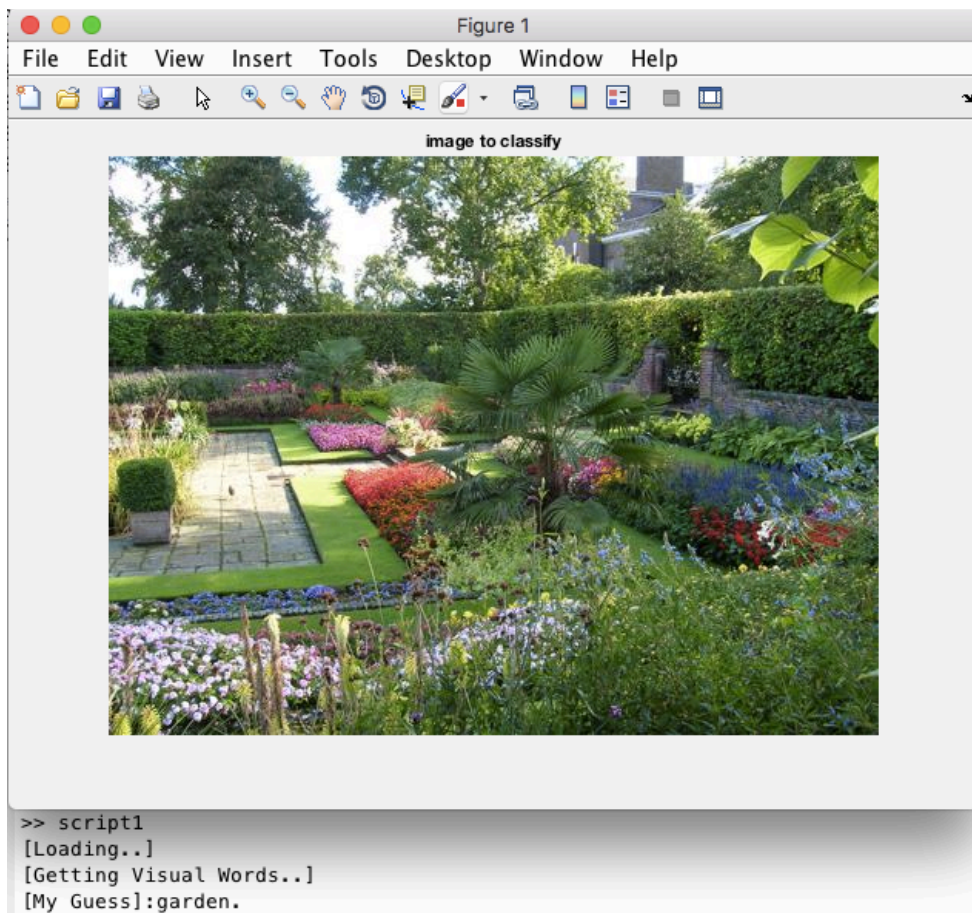
Q2.5 Report the confusion matrix and accuracy for your results in your write-up.

conf =

```
7   2   0   3   7   0   1   0
2   7   0   3   7   1   0   0
0   1  18   0   1   0   0   0
2   1   0  15   1   0   0   1
4   6   0   2   8   0   0   0
0   1   1   0   0  11   6   1
0   1   0   0   0   3  16   0
1   4   4   3   2   1   4   1
```

accuracy = 51.8750 %

	Art gallery	Computer room	Garden	Ice skating	Library	Mountain	Ocean	Tennis Court
Art gallery	7	2	0	3	7	0	1	0
Computer room	2	7	0	3	7	1	0	0
Garden	0	1	18	0	1	0	0	0
Ice skating	2	1	0	15	1	0	0	1
Library	4	6	0	2	8	0	0	0
Mountain	0	1	1	0	0	11	6	1
Ocean	0	1	0	0	0	3	16	0
Tennis Court	1	4	4	3	2	1	4	1

There are some classes/samples that are more difficult to classify than the rest using the bags-of-words approach. As a result, they are classified incorrectly into other categories.

Q2.6 (5 points) List some of these classes/samples and discuss why they are more difficult in your write-up.

The samples like garden and ocean enjoys more accuracy as compared to art gallery and library do. This can be attributed to more distinct colors and textures and thus more distinct clusters formed for the garden (green) and the ocean(blue) when spatial pyramid matching is used. Meanwhile, the library, computer room and the art gallery share common characteristics like a closed room and walls because of which the clustering gets wrong on some occasions.

To briefly state:
Samples of - **Art Gallery**, **Computer Room**, **Library** and **Tennis Court** samples are more difficult to classify than the rest using the bags-of-words approach.

Q2.7 The performance is expected to improve ever so slightly with a more value of random sampling points or alpha. Also increasing the number of clusters would enable to capture all the available colors in the image. To improve performance more training data can be used. Coupled with that, if the training data is in different view angles, that could train the system better. Also different filters could be used to improve the performance. Some of them are like the Kalman filter which is an optimal estimation algorithm that uses techniques of capturing images/scenes from multiple angles and progressions.