16-720 Computer Vision: Homework 1 (Spring 2020)
Spatial Pyramid Matching for Scene Classification

PROBLEM1.1.1

Three broad categories of filters used are:

1) **Gaussian Filter:**
    - It is a low pass filter effective to reduce noise in images. It is used to collect noiseless features from images having noise.
    - It provides more weight to the center pixel due to which, not much spatial information is lost. As it has weighted response, it is better than mean or median filters.
    - It has similarities with human perception system
2) **Laplacian of Gaussian Filter:**
    - It is a second derivative filter which gives zero response for uniform regions in an image and positive/negative response in regions where there are changes.
    - These filters are commonly used to detect blob-like features in images.
    - The Laplacian filter is usually applied after a Gaussian filter to smooth it and improve the response, as derivative filters are highly sensitive to noise.
3) **First Derivative Filter:**
    - These filters are applied in the X and Y directions separately, and are used to detect vertical and horizontal edge-like features respectively.
    - The derivative filters are highly sensitive to noise, Gaussian filters are used to smooth it and improve the response.

Use of Multiple Scales:
Multiple scales of filter responses are required because the size of features may vary in different images, also we would like the filter bank to pick all these multi-sized features. For example, in order to detect an edge which is far away from camera we may need filter with smaller kernel, similarly, to detect edges from larger sized objects we need larger kernel filters. This can be achieved by varying the size of kernels by changing the sigma (scale) value.

PROBLEM 1.1.2

The collage of the images obtained from different filters are given below.

Figure 1 Kitchen

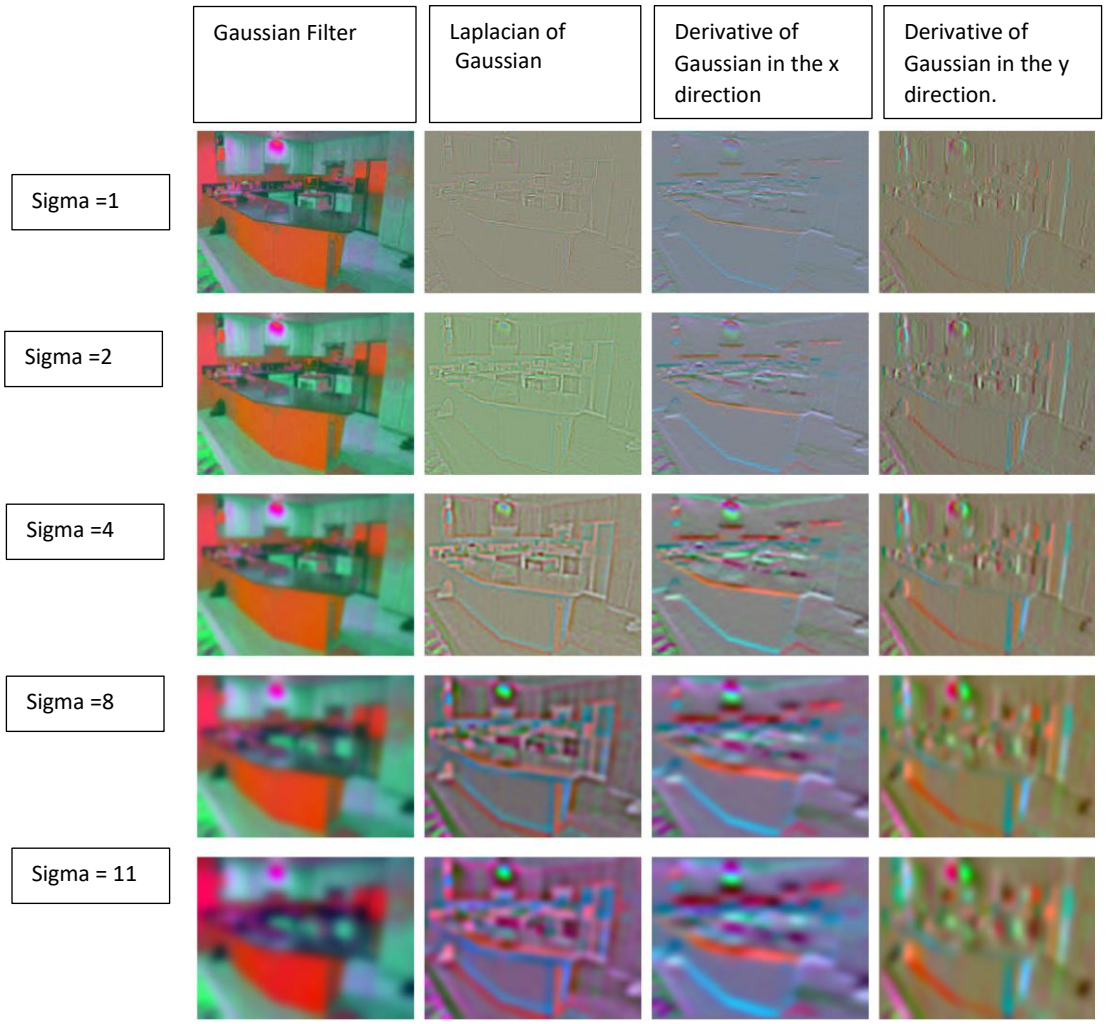
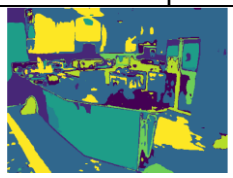
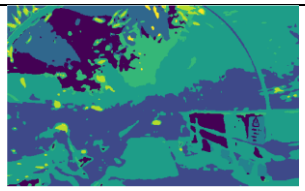


Figure 2 Collage of filters

PROBLEM 1.3

Images of kitchen, aquarium and windmill and their wordmaps with default values of hyperparameters are given below. From the image we can see that for well- structured images like the kitchen and windmill images, the wordmaps indicate the features more clearly and boundaries are well defined. However, for the images like aquarium, less characteristics can be differentiated from the wordmap and the boundaries are not as clearly defined.

Table 1 Comparison of Original Image and Wordmap

Original Image	Wordmap
 Kitchen	
 Aquarium	
 Windmill	

For dictionary generation K = 100 and alpha=150 was chosen after tuning in parameters. These pair worked appropriately when evaluating the recognition system. Images of which are shown below. It can be seen that the classification of the image features is done more accurately using the tuned hyperparameters

Figure 3 Tuned images and its wordmaps

PROBLEM 2.5

1) Using default Hyperparameter
Filter scale- [1,2]
K-10
Alpha-25
L-1
Confusion Matrix-
[[30. 3. 1. 2. 1. 0. 10. 4.]
[3. 15. 6. 6. 1. 0. 1. 5.]
[0. 6. 22. 1. 2. 3. 2. 4.]
[4. 14. 1. 23. 14. 0. 0. 0.]
[1. 6. 5. 9. 21. 4. 7. 4.]
[1. 1. 1. 3. 3. 29. 7. 7.]
[6. 2. 5. 5. 7. 10. 19. 9.]
[5. 3. 9. 1. 1. 4. 4. 17.]]

Accuracy-0.44 (44%)

2) Using tuned Hyperparameter
Filter scale- [1,2,4,8,11]
K-100
Alpha-150
L-3
Confusion Matrix-
[[40. 0. 1. 2. 1. 2. 5. 1.]
[0. 31. 4. 2. 1. 0. 1. 3.]
[0. 5. 26. 1. 1. 2. 1. 8.]
[1. 3. 0. 34. 14. 1. 0. 0.]
[3. 3. 1. 9. 27. 1. 5. 0.]
[2. 3. 3. 0. 2. 39. 6. 4.]
[3. 0. 1. 1. 3. 0. 28. 1.]
[1. 5. 14. 1. 1. 5. 4. 33.]]

Accuracy-0.645 (64.5%)

PROBLEM 2.6

From the confusion matrix above, following can be inferred:

1) Misclassification of categories is observed due to common features in them.

Table 2 Misclassification examples

Classes in Misclassification	Image1	Image2	Comments
Highway and windmill	 Highway	 Windmill	Both of the classes have roads in common
Kitchen and Laundromat	 Kitchen	 Laundromat	Both classes have similar equipment. Due to which recognition system were unable to differentiate
Park and Waterfall	 Park	 Waterfall	Presence of water in both of them.

2) The best classification is done between aquarium and desert it has very few features in common, recognition system is able to distinguish it from the other classes.

Figure 4 Aquarium

Figure 5 Dessert

PROBLEM 3.1

Table 3 Accuracy for different hyperparameters

Run No	Filter_scales	K	alpha	L	Accuracy %
1	[1,2,4,8,11]	25	150	3	62.25
2	[1,2,4,8,11]	100	25	3	63.25
3	[1,2,4,8,11]	100	150	3	64.5

It was observed that by using default values of the hyperparameters accuracy of **44%** was achieved.

1) **Run1**

Filter Scales were increased from [1,2] to [1,2,4,8,11]

K was increased from 10 to 25

Alpha was increased from 25 to 150

L was increased from 1 to 3

The combined effect of all the changes caused the accuracy of the model to increase to 62.25%

Effect of increase in the hyperparameters:

Increase in K: As the value of K increases the elements in a cluster reduces. So average distortion will decrease due to which classification accuracy increases.

Increase in alpha: The increase in the alpha value provides more data to the classification algorithm thus improving accuracy.

Increase in L: With the increase in the number of layers more details can be captured from images which will help to classify the images better.

Increase in filter_scales: Multiple scales of filter responses will help to pick all the multi-sized features and improve accuracy

2) **Run2**

In this run the value of K was increased further but the value of alpha was reduced to default value. From the results it can be inferred that the improvement in K will affect the accuracy more than alpha.

3) **Run3**

In this run both the values of K and alpha was increased which gave the maximum accuracy value of 64.5%

PROBLEM 3.2 (Extra Credit)

Hyperparameter Values Used:
Filter scale- [1,2,4,8,11]
K-100
Alpha-150
L-3
Confusion Matrix:
[[39. 1. 1. 1. 0. 2. 2. 0.]
[0. 29. 4. 1. 1. 0. 1. 3.]
[1. 5. 28. 1. 1. 2. 1. 8.]
[1. 3. 0. 33. 12. 1. 0. 0.]
[3. 4. 1. 12. 29. 2. 5. 1.]
[2. 4. 3. 0. 2. 40. 7. 5.]
[2. 0. 1. 0. 4. 0. 32. 0.]
[2. 4. 12. 2. 1. 3. 2. 33.]]

Improved Accuracy: 0.6575 (65.75%)

1) **What you did?**
   The hyperparameter values which gave maximum value of accuracy were maintained. **The weights of SPM were changed for each layer.** The weights given earlier were set as 2^{-L} for layers 0 and 1, and $2^{(l-L-1)}$ for rest of the layers. When number of layers were 3, weights were 1/4, 1/4 and 1/2 for 0,1 and 2 layers respectively. For the same three layers the weights were reduced to 1/8, 1/8 and 1/4 for layers 0,1 and 2 respectively.
2) **What you expected would happen?**
   As the weights give priority to the histogram with respect to the layer number, it was seen that on reducing the weights for each layer by the same scale the accuracy of image recognition will not be affected.
3) **What actually happened?**
   The accuracy of image recognition improved by reducing weights.