CSE 573: Computer Vision and Image Processing: Homework 1
Spatial Pyramid Matching for Scene Classification

Last Name: Nanwani

First Name: Charushi

UBIT ID: charushi

Person #: 50248736

Q 1.0

1. Gaussian Filter: Gaussian filter or Gaussian smoothing is a 2-D filter which is used to blur out images by removing any noise and detail. The degree of smoothing is decided by the standard deviation of Gaussian derivation.
   The Gaussian outputs a `weighted average' of each pixel's neighborhood, with the average weighted more towards the value of the central pixels. This is in contrast to the mean filter's uniformly weighted average. Because of this, a Gaussian provides gentler smoothing and preserves edges better than a similarly sized mean filter.

2. LoG Filter: The Laplacian of Gaussian or Log filter is a measure of spatial derivative of image. It is used to highlight regions of sudden intensity changes. Thus, it is used for edge detection. This filter is generally applied to an image which is already a filtered and smoothed output of Gaussian filter.
   At a reasonably sharp edge between two regions of uniform but different intensities, the LoG response will be:
   - zero at a long distance from the edge,
   - positive just to one side of the edge,
   - negative just to the other side of the edge,
   - zero at some point in between, on the edge itself.
3. Dx and dy scales: The Gaussian filter can be separated into x and y components. Thus the convolution can be performed by first convolving in the x direction separately and then convolving with another filter in the y direction.

   Reference:
- http://homepages.inf.ed.ac.uk/rbf/HIPR2/log.htm
- http://homepages.inf.ed.ac.uk/rbf/HIPR2/gsmooth.htm

Q 1.1

Image used: **'\art_gallery\sun_akdxvnqeibphzzfu.jpg'**
Montage: **'\art_gallery\sun_akdxvnqeibphzzfu_montage.fig**'

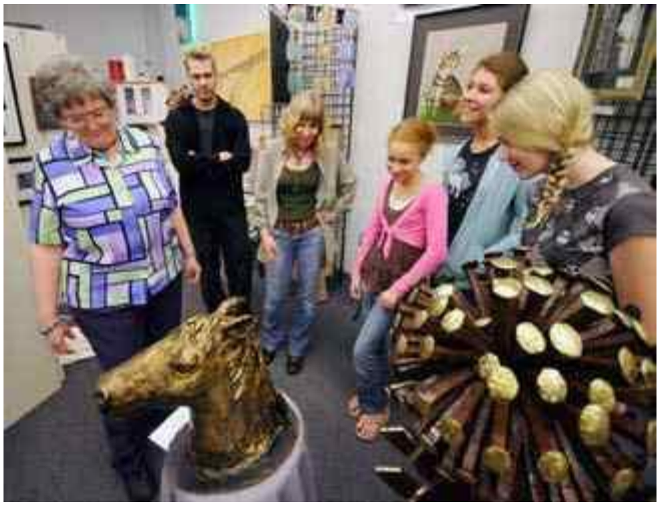

Figure 1 1.1 art_gallery\sun_akdxvnqeibphzzfu.jpg

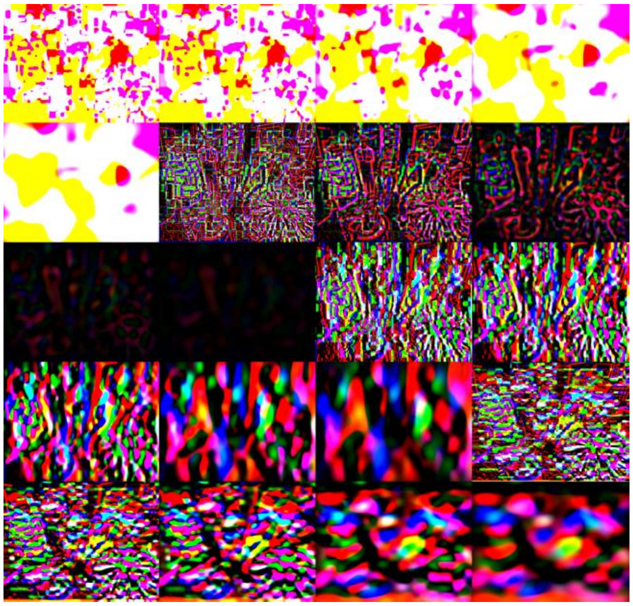

Figure 1 1.1 art_gallery\sun_akdxvnqeibphzzfu.mat

Q 1.3

1. Image: '**art_gallery\sun_akdxvnqeibphzzfu.jpg**'
   WordMap: '**art_gallery\sun_akdxvnqeibphzzfu.mat**'

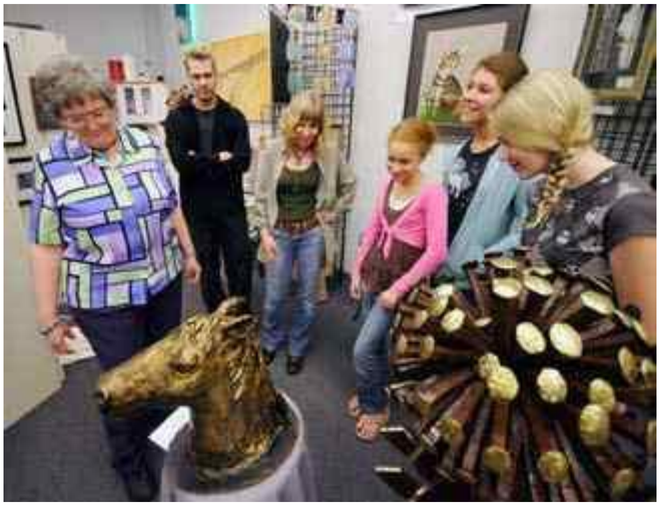

Figure 1 1.3 art_gallery\sun_akdxvnqeibphzzfu.jpg

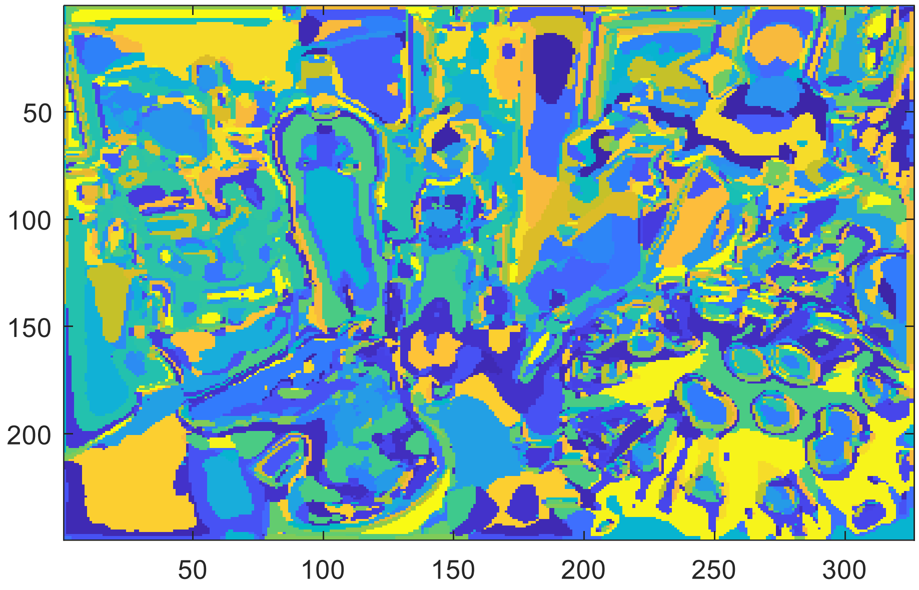

Figure 2 1.3 art_gallery\sun_akdxvnqeibphzzfu.mat

2. Image: '**art_gallery\sun_asnpcwugrpbvfznx.jpg**'
   WordMap: '**art_gallery\sun_asnpcwugrpbvfznx.mat**'

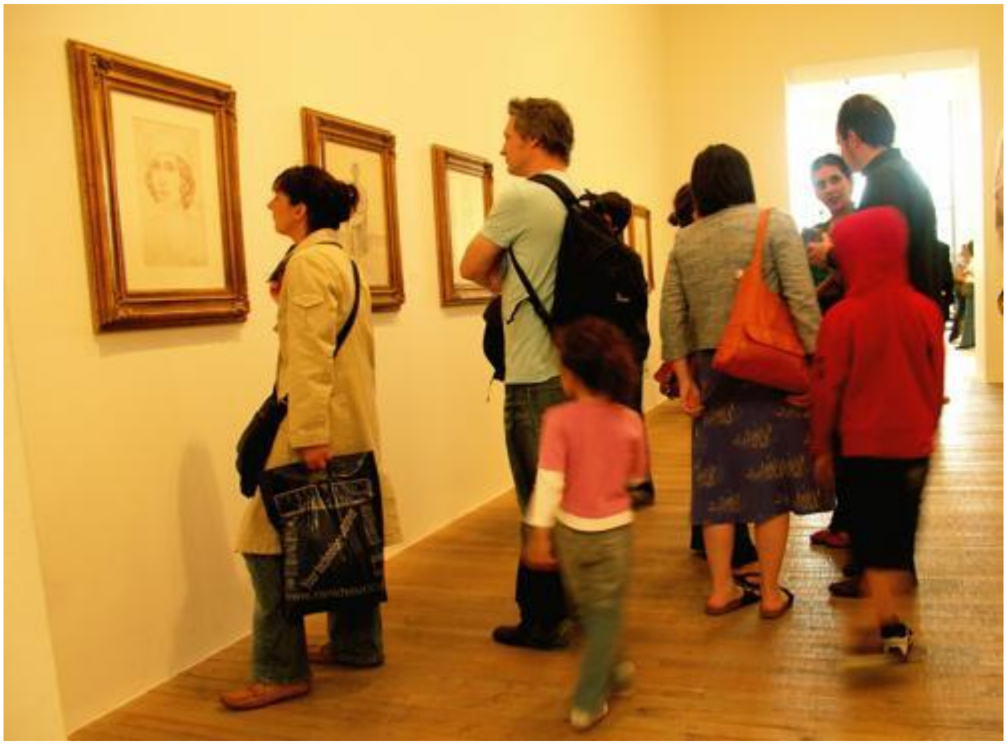

Figure 3 1.3 art_gallery\sun_asnpcwugrpbvfznx.jpg

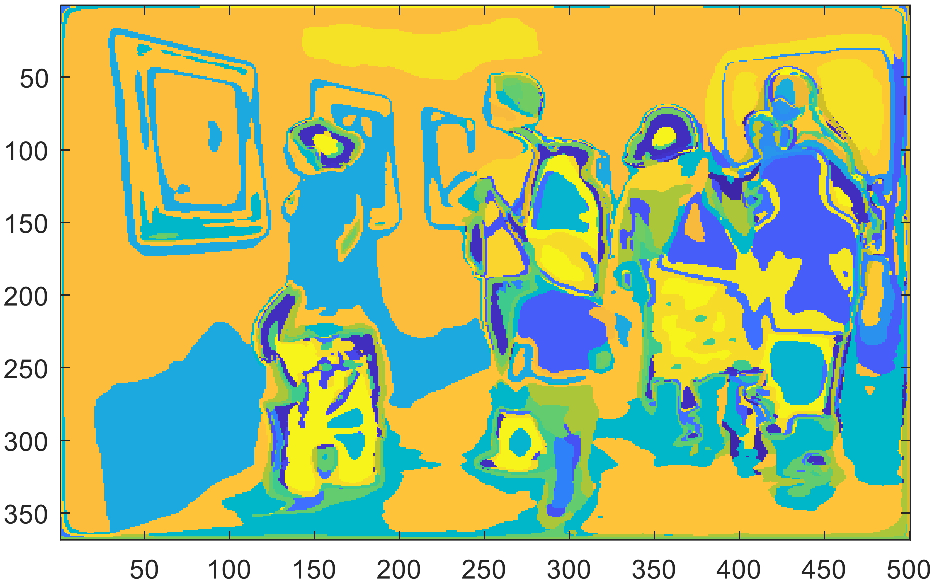

Figure 4 1.3 art_gallery\sun_asnpcwugrpbvfznx.mat

3. Image: ‘art_gallery\sun_acoitvzbpriogtkp.jpg’
   WordMap: ‘art_gallery\sun_acoitvzbpriogtkp.mat’

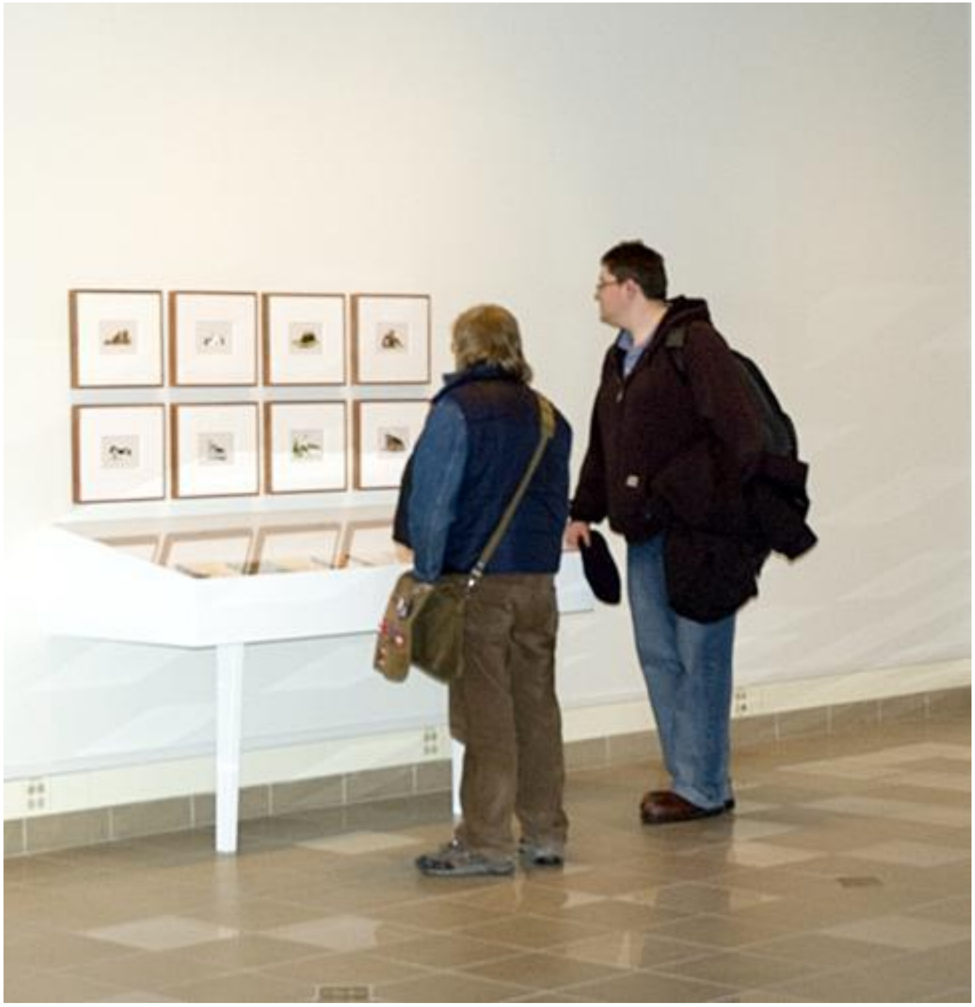

Figure 5 1.3 art_gallery\sun_acoitvzbpriogtkp.jpg

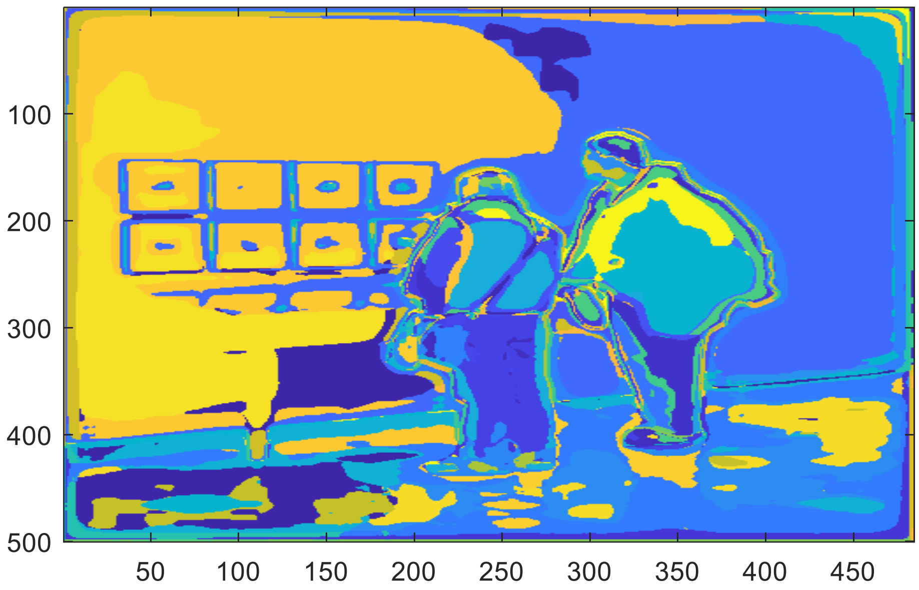

Figure 6 1.3 art_gallery\sun_acoitvzbpriogtkp.mat

Q 2.5

Confusion Matrix

Confusion Matrix	Art_gallery	Computer_room	Garden	Ice_skating	Library	Mountain	Ocean	Tennis_court
Art_gallery	6	2	0	5	6	0	0	1
Computer_room	3	5	0	3	9	0	0	0
Garden	0	0	19	0	0	0	0	1
Ice_skating	1	3	2	14	0	0	0	0
Library	4	5	0	2	9	0	0	0
Mountain	0	1	3	0	0	12	3	1
Ocean	0	0	0	1	0	4	14	1
Tennis_court	1	3	5	2	2	1	0	6

Accuracy

53.1250 %

Q 2.6

The accuracy of this code is 53.125%. The classes ‘**computer_room**’, ‘**art_gallery**’, ‘**tennis_court**’ were more difficult to classify and map. This is because the edge detection and intensity changes could not be differentiated easily.

For ‘**art_gallery**’, color scheme and the intensity changes in few of the images are similar to that of library and ice skating. So the filter response obtained might be similar to that of an image which belonged to ‘**ice_skating**’ class or ‘**library’**.

For ‘**computer_room**’, most of the images were wrongly classified as ‘**library’**. This is because, in many images, the structure and placement of various objects in the room is somewhat that of a library, which makes the differentiation harder.

Also, for ‘**tennis_court’**, since most of the courts are green, this produced the same response as that of a garden image. Thus, images actually belonging to ‘**tennis_court**’ were classified as ‘**garden’.**

Also, the training set used in the example (1349) is quite small. So, this could also have affected the classification. The classification is also decided by the values of a and K. A small value of a(alpha) results in considering only a few random pixels (which might leave out some pixels having good intensity).
Also a small value of k will result in a few clusters, thus grouping different parts. This might cause an image to be classified into a different class.

Q 2.7

After increasing K from 120 to 200 and keeping a same (120), the accuracy increases from 53.125 to 53.75 %. The confusion matrix for K=200 and a=120 is given below.

Confusion Matrix	Art_gallery	Computer_room	Garden	Ice_skating	Library	Mountain	Ocean	Tennis_court
Art_gallery	8	1	0	4	6	0	1	0
Computer_room	1	6	0	1	12	0	0	0
Garden	0	2	16	0	0	1	1	0
Ice_skating	1	2	0	15	2	0	0	0
Library	3	8	0	0	8	1	0	0
Mountain	3	1	2	2	2	8	2	0
Ocean	0	0	0	0	1	2	17	0
Tennis_court	1	3	5	0	3	0	0	8

Accurracy:

53.7500 %

If we increase the value of a up to 150, we can get an accuracy of 55%. But this takes a computation time of around 60 minutes.

Thus, by increasing value of a and K accordingly, we can improve the accuracy of the above code.