CSE 573: Computer Vision and Image Processing

Spatial Pyramid Matching for Scene Classification

Name: Prachi Jayakumar Patil

Person Number: **50247309**

UBIT Name: **ppatil3**

---

1. Filters used in the filter bank:

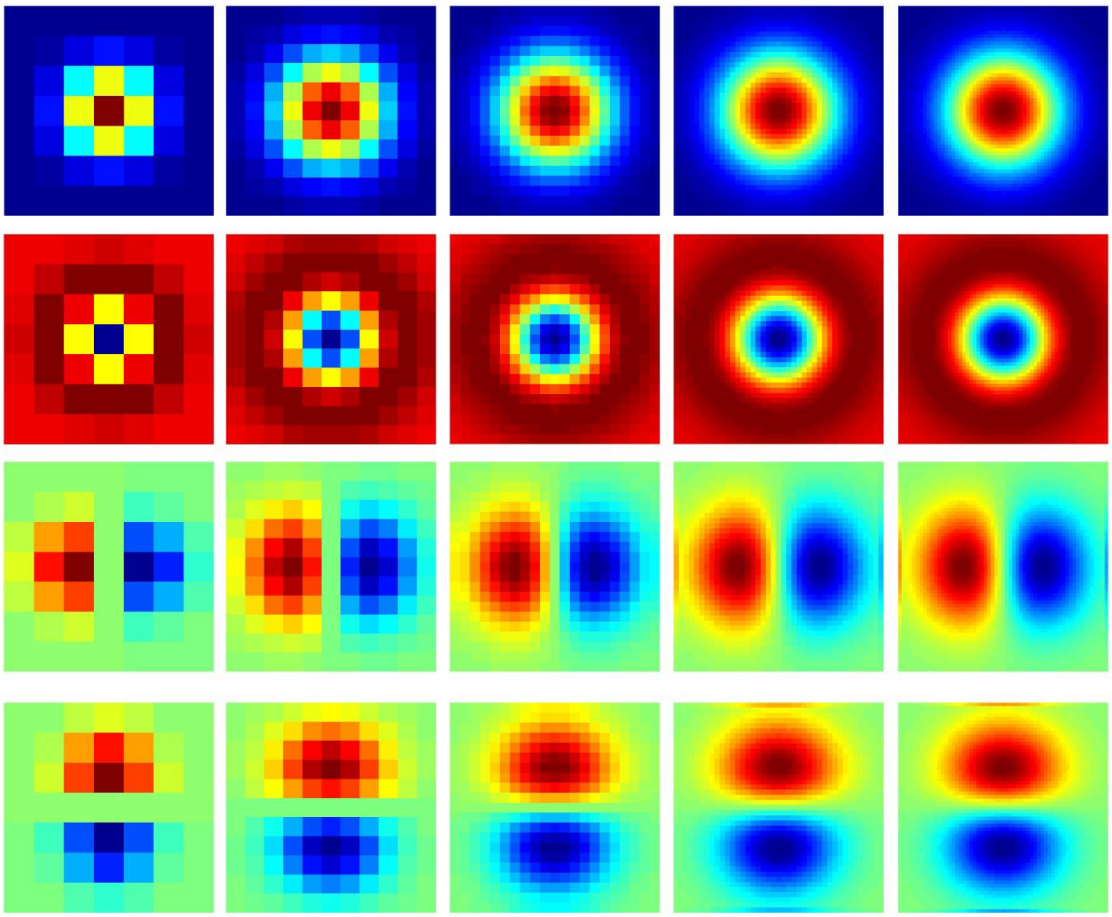

Figure 1: Multiscale Filter Bank

1) **Gaussian Filter:**
   Gaussian filter a is low-pass filter. It suppresses high-frequency detail(noise), while preserving the low-frequency parts of the image. It basically blurs the image.

2) **Laplacian Filter:**
   The Laplacian is used for edge detection as it detects the regions of high frequency changes.

3) **Horizontal Scale:**
   The horizontal scale checks for the vertical edges.

   Filter: [-1 0 1]

4) **Vertical Scale:**

   The vertical scale checks for the horizontal edges.

   Filter: [-1 0 1]'

   Gaussian filters:
   a) Gaussian low pass filter
   b) Horizontal Scale filter
   c) Vertical Scale filter

2. FilterResponse

Size : M x N x 3F

After application of 20 filters on the sample image(fig a), an image collage(fig b) is generated.

Figure 2a Sample Image

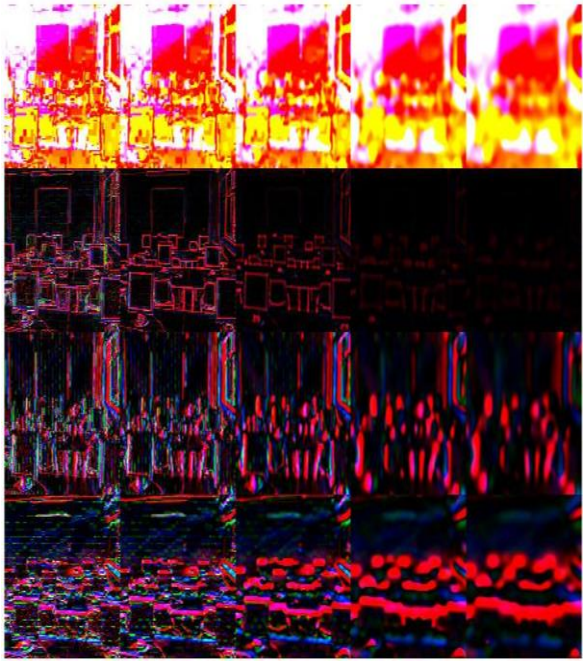

Figure 2b Collage Image

While applying the filters, the imfilter function is used for convolving the filters in filter bank with image. The filtering option mentioned in the function are:

'conv': performs multidimensional filtering using convolution.

'same': to keep the output size as the input size

'replicate': in this case, the values of the pixel outside the image are determined by replicating its neighboring pixel value. The zero padding can cause the dark bands around the edges, whereas the replicate argument will eliminate this effect.

1.2

Alpha = 100

K = 150

3. WordMap Image

a)

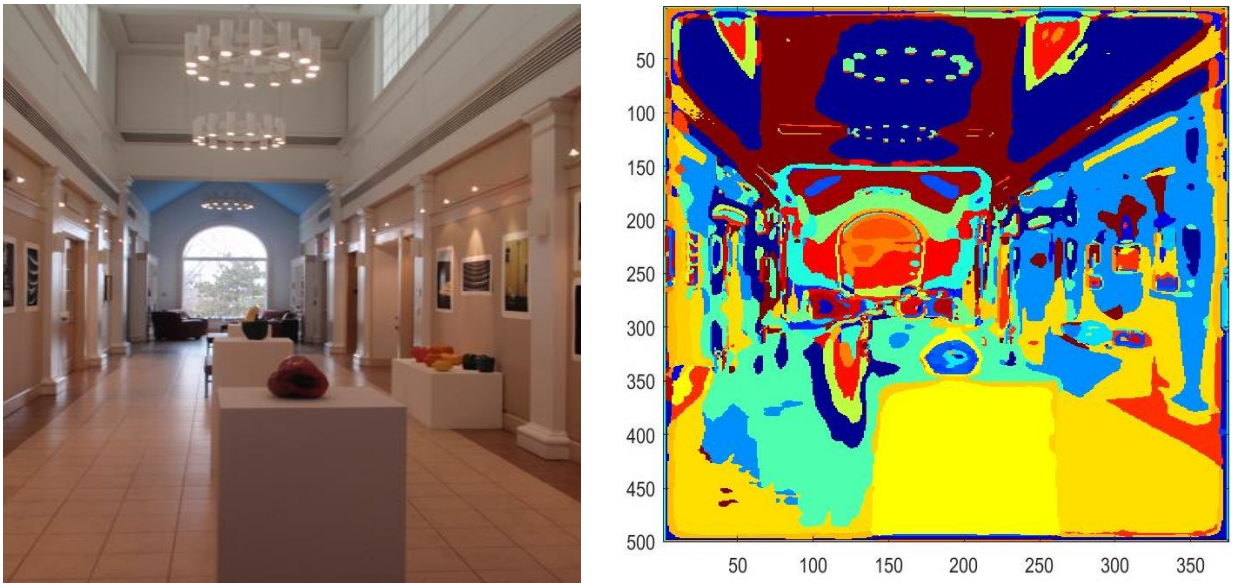

Figure 3: a) Original Image b) WordMap Image

b)

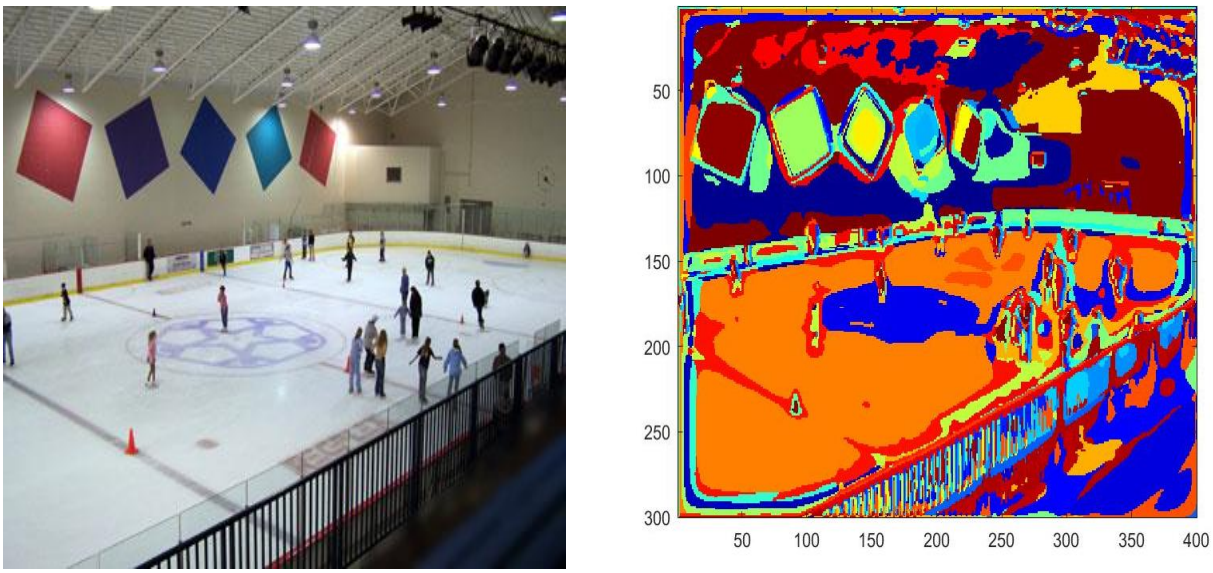

Figure 4: a) Original Image b) wordMap Image

c)

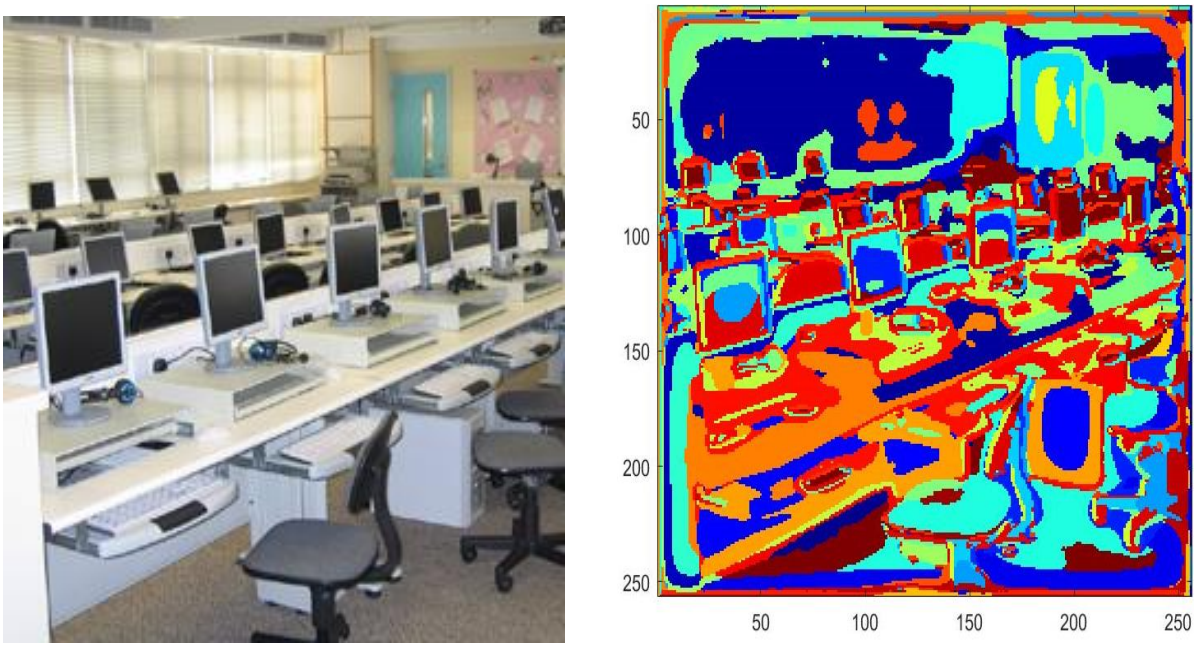

Figure 5: a) Original Image b) WordMap Image

d)

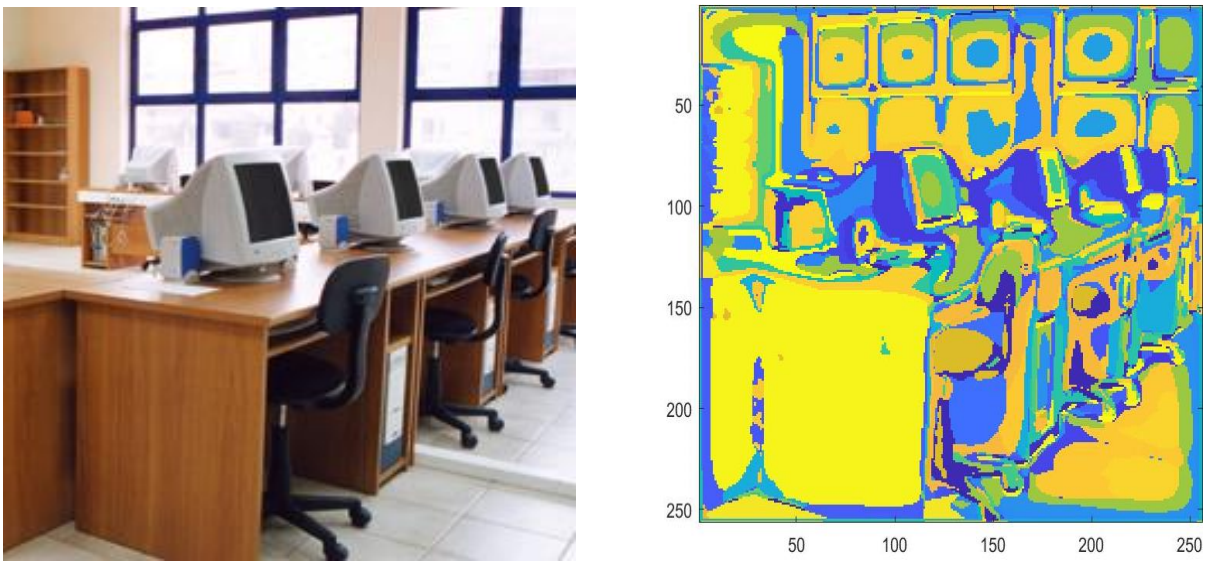

Figure 6: a) Original Image b) wordMap Image

e)

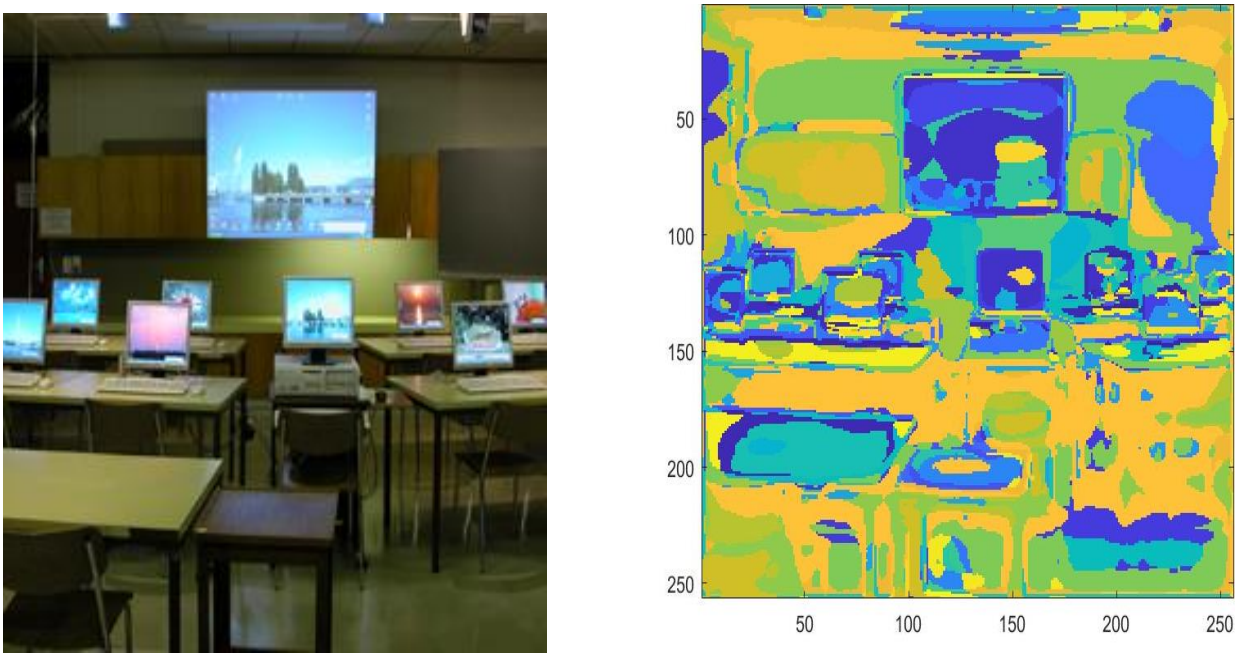

Figure 7: a) Original Image b) wordMap Image

The wordMap images define the shapes of the object. The similar colors/shapes in the image are classified into similar categories or into bag of words. For the category, computer room, we can see that the computer shape is defined in the wordMap and we can identify from its wordMap that it belongs to computer room.

4. Results

Confusion Matrix:

7	2	0	3	6	1	1	0
4	3	0	4	6	3	0	0
0	0	17	0	2	1	0	0
1	1	2	14	2	0	0	0
6	8	0	1	5	0	0	0
1	1	0	0	1	6	11	0
0	0	0	0	0	5	15	0
1	3	3	2	3	1	2	5

Accuracy: 45%

5. Inference

From the confusion matrix we can infer the following,

a) The accuracy for the category ‘garden’ is the highest (). This is because the color scheme and the variety of the visual words in the garden category are almost similar. So it is easy to identify the garden images.
b) Similarly, for the ‘ice skating’ is high. As the white color is predominant. This helps in identifying the ice skating category images
c) The accuracy for the category ‘tennis court’ is the least (). This is because there are less objects in the tennis court to define its properties.
d) The below tennis court image is classified as ocean. This is because the ocean is classified by sea and the sky. In the below image, the glass is classified as the sky and the bottom part is considered as water. So the system classified it as ocean instead of tennis court.

e) The below computer room is classified as ‘mountain’. This is because the system is not able to identify the computer like structures. The corresponding images of the mountain shown below is considered and it is classified as mountains instead of computer room.