

Homework-1

N Dinesh Reddy
dnarapur@andrew.cmu.edu

February 6, 2017

1 Representing the world with visual worlds

Problem 1.0. The filters help to convert an image to a meaningful representation. We elaborate on properties captured by each filter below:

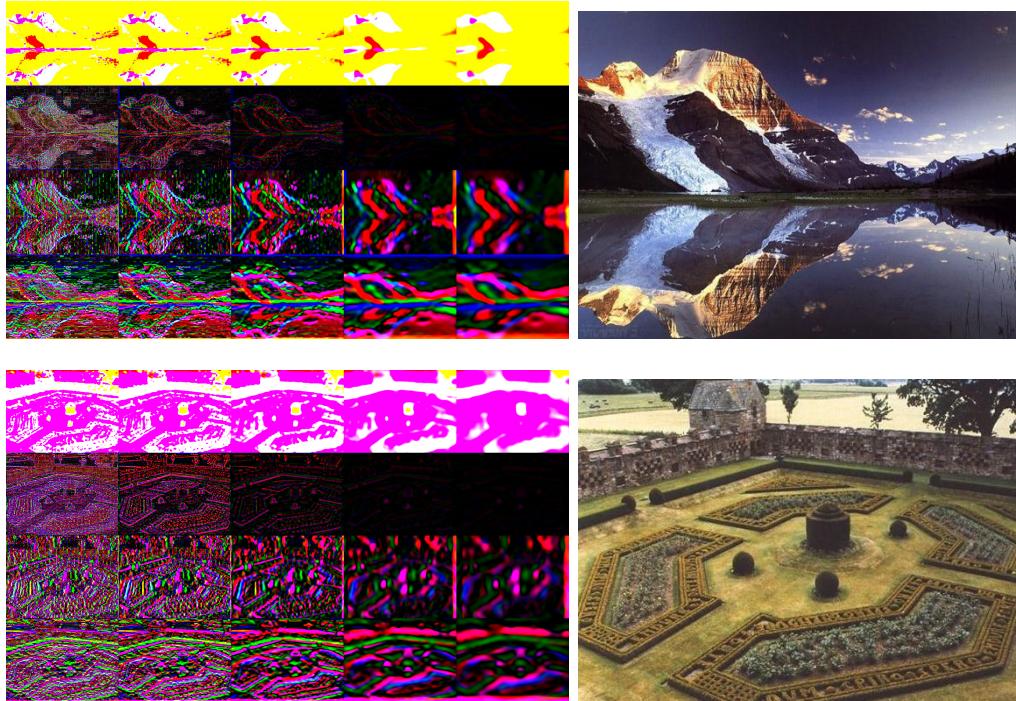
Filters 1-5: The displayed filters are **gaussian** filters. These filters help in capturing the information in the surrounding pixels. They are a weighted sum of the intensities of the surrounding pixels with a decreasing weight based on the distance from the center pixel. These filters extract the color information of the neighbourhood into the pixel.

Filters 6-10: These filters represent the **laplacian of Gaussian** filters. These filters are used to capture rapid changes in the intensity of color of the image. These filters belong to the derivative class of filters. At a sharp edge between two regions, the response will be zero far away from the edge and on the the edge, positive to one side and negative on the other side. This filter gives the highest value to corners and decreasing value to lighter edges.

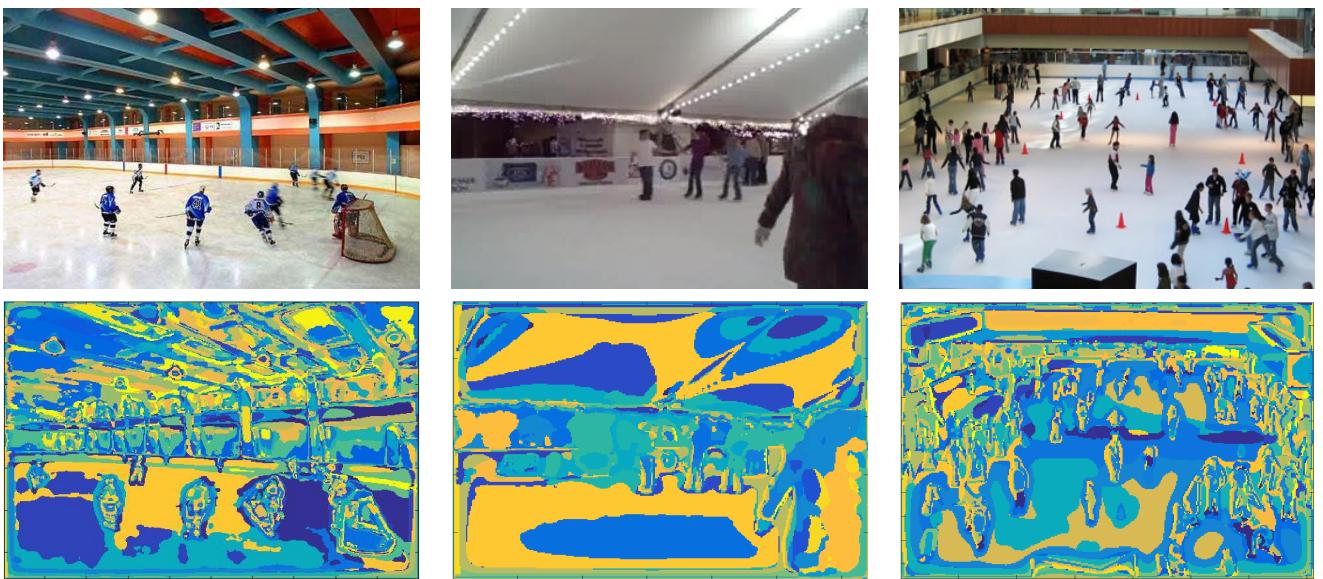
Filters 11-15: These filters can be named as directional filters. They are produced by apply a directional bias to a Gaussian filter. The following filters capture variation in appearance in horizontal direction. The filter gives higher response when left neighbourhood posses higher intensify to the right neighbourhood of the center pixel. This can be attributed because of the mask having negative values on the right hand side of the image.

Filter 16-20: The filters in this category are very similar to the earlier filter category but with the variation that they are vertical filters. They capture the variation in the vertical direction and give higher response when pixels in the upper region have higher value to the lower pixels.

Problem 1.1. We now apply the filters to an LAB image and display the results for two images below:



Problem 1.3. We compute the wordmap and display the results for 3 images from the icecate category of the dataset. For all the experiments we have used $K = 200$ and alpha = 100:



	art gallery	computer	garden	ice skating	library	mountain	ocean	tennis
art gallery	11	0	0	4	4	0	0	1
computer	3	8	0	6	2	1	0	0
garden	0	0	19	0	1	0	0	0
ice skating	0	3	0	16	0	0	0	1
library	6	5	0	1	8	0	0	0
mountain	1	1	5	0	2	9	2	0
ocean	2	2	0	2	1	2	8	3
tennis	1	5	1	2	1	2	0	8

Table 1: 'Confusion matrix for the sample data'

Problem 2.5. The accuracy of the algorithm comes out to be 54.37% when the K=200 and alpha = 100 is taking to compute the visual words. the confusion matrix is displayed in table 1 for all the classes.

Problem 2.6. Some of the classes which are tough to classify using the bag-of-words implementation areas can be seen from the confusion matrix are:

- a) computer room
- b) library
- c) ice skating

lets try to discuss each failure case. The computer room has a lot of clutter in the image for it to be classified into a specific indoor category. As can be seen from the confusion matrix most of the computer room test images are classified as ice skating or library. This is because of the visual words computed for computer has wide variation in color and locations. This has resulted in classification of the computer lass across all other classes. This justifies the fact that a computer room has too much variation for a easy classification compared to the other classes.

The library class faces similar problems to the computer room class, where the structure of the indoor scene and wide variation in image colors is causing the miss classification of the library to other similar classes like art gallery and room. To improve the accuracy for such situations of indoor scenes, one of the method can be to increase the number of pyramid levels. The intuition behind this is, scenes like computer room have larger objects and lesser variation within a smaller window compared to art gallery and library which contain smaller objects.

The other class which we are seeing a high misclassification accuracy is for the ice skating category. As can be seen from the confusion matrix the ice skating class has wide spread distribution of wrong classification into nearly all classes. this can be attributed to the fact that among the dataset this class is the only category which has both a uniform plane surface with cluttered humans skating in the field. This leads to the wrong classification of this category into both cluttered environments like computer room, art gallery and classes with uniform texture like tennis court and ocean.

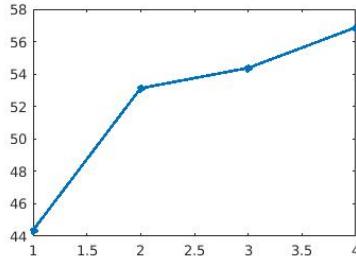


Figure 1: pyramid level vs accuracy plot

	art gallery	computer	garden	ice skating	library	mountain	ocean	tennis
art gallery	9	2	0	3	5	0	0	1
computer	3	13	0	3	1	0	0	0
garden	0	0	18	0	0	1	0	1
ice skating	3	0	0	15	1	0	1	0
library	1	4	0	1	14	0	0	0
mountain	0	1	2	0	1	12	2	2
ocean	0	1	0	0	1	3	14	1
tennis	1	1	3	0	1	1	2	11

Table 2: 'Confusion matrix using neural network'

Problem 2.7. To look at improving the accuracy of the current algorithm the first set of experiments conducted are computing accuracy versus the number of pyramid levels. As shown in the figure 1 we fix all other parameters for the algorithm and vary the number of pyramid levels in the spacial matching algorithm as expected with increase in the spacial subdivisions the accuracy of the algorithm improves. We get an accuracy of 56.8750% for pyramid level of 4.

To improve the accuracy the second improvement we propose is to apply neural network for the classification task keeping the same feature extraction algorithm. We show the accuracy boast upto 66.25% using this neural network with the same features used using the histogram kernel matching. We can further improve the accuracy by using more intelligent feature extraction methods and by preprocessing the images. The confusion matrix after applying a two layer fully connected network is shown in table 2. As you can see the confusion matrix clearly shows an improvement in all the classes and does not have any bias towards specific classes. This can be attributed to the fact that a fully connected layer over the histogram nullifies any local effected which were causing miss classification earlier. We have a used a simple one layer fully connected network with 32 hidden layer values. We have experimented with different layer architectures and have observed over fitting with multiple layers and misclassification with lee hidden layer values. Further improvement in feature extraction schema can improve the accuracy of the algorithm. All the experiments are conducted with K=200 and alpha=100. Experimented varying these parameters to find variation 2-4 percent in accuracy.