Computer Vision 16-720 HW1

Xiaoyu Bai (xiaoyub)

Q1.0

The first five filter banks are Gaussian filter. Gaussian filter is used to `blur' images and remove detail and noise. In this sense it is similar to the mean filter, but it uses a different kernel that represents the shape of a bell-shaped hump. This kernel has some special properties which are detailed below. [1]

The second five filter banks are Laplacian of Gaussian filter, meaning to smooth the image first by a convolution with a Gaussian kernel and then apply Laplacian filter. And Laplacian is a derivative operator; its uses highlight gray level discontinuities in an image and try to deemphasize regions with slowly varying gray levels. [2][3]

The third five filters are dX scale Gaussian filter which smooths the image along x direction.

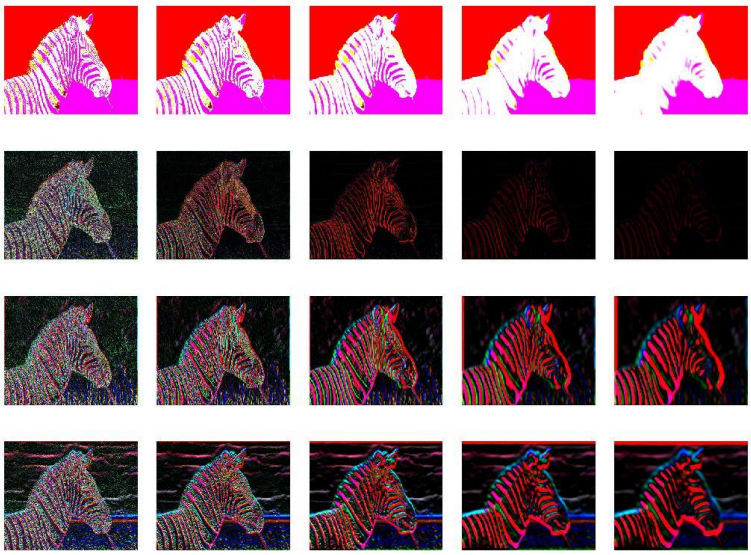
The last five filters are dY scale Gaussian filter which smooths the image along y direction.

Q1.1

I use the zebra image from HW0 as an example.

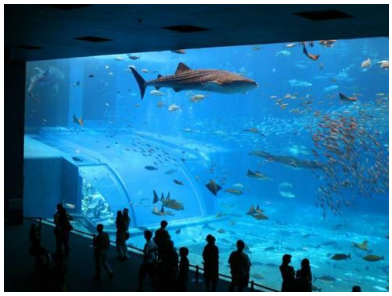
Original image:

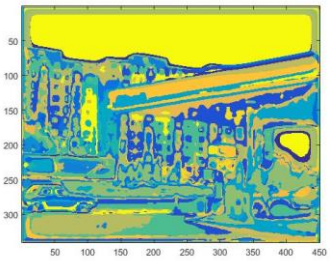
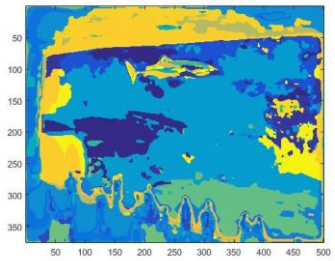
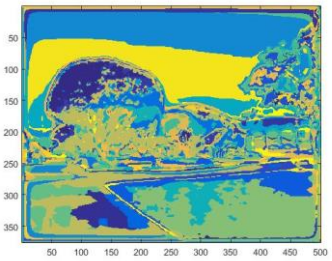
Collage of 20 filter responses:



Q1.3

Heat map of 3 sample images:



Since we use the nearest vocabulary to represent each pixel, the figures will eliminate some of the minor difference between similar pixels. By omitting those details, we can use the new representation to capture the primary information (mainly color and shape) of each figure.

Q2.5

My confusion matrix is as below:

16	1	1	0	0	0	2	0
0	10	1	1	5	0	2	1
1	1	13	2	0	1	0	2
1	0	2	14	0	0	1	2
0	2	0	3	13	0	2	0
0	1	0	1	2	16	0	0
0	0	3	1	4	0	9	3
1	1	2	0	2	8	2	4

The results are after optimization. Here I use α=100 (Number of elements randomly picked in filter responses) and K=200 (dictionary size).

The accuracy is 59.375%.

Q2.6

Here we analyze some failed cases. From my understanding, the bag-of-words approach mainly focuses on small patches of images. If some small patches look close to a certain class which it does not belong to, then it is highly possible to be misclassified. For example, this image from aquarium has been classified as bridge because it has curves which look just like bridge:

And this image of parking lot has been classified as waterfall and I guess it is because the shape of the trees in background look like waterfall:

Reference:

[1] http://homepages.inf.ed.ac.uk/rbf/HIPR2/gsmooth.htm

[2] http://www.tutorialspoint.com/dip/laplacian_operator.htm

[3] http://fourier.eng.hmc.edu/e161/lectures/gradient/node8.html