HW1 Spatial Pyramid Matching for Scene Classification

Yasser Corzo

September 2023

Q1.1.1

Properties that Gaussian filters pick up are areas where there are bright pixels. Since the Gaussian outputs a "weighted average" of each pixel's neighborhood (i.e. the pixels that are covered by the kernel), the average is weighted more towards the value of the center pixels. As a result of this, areas in the image where pixels have a high value are "weighted" more. In addition, since the center pixels are weighted more, Gaussian filters provide gentler smoothing and remove noise. Properties that Laplacian of Gaussian filters pick up are edges. Properties that the derivative of Gaussian in the x direction pick up are vertical edges. Properties that the derivative of Gaussian in the y direction pick up are horizontal edges. Multiple scales of filter responses are needed because it allows us to capture more information at different levels (since increasing the scale can blur the image more, etc) by preserving edges and reducing noise. Hence, more features can be captured which will be useful for creating a dictionary of visual words from the filter responses.

Q1.1.2

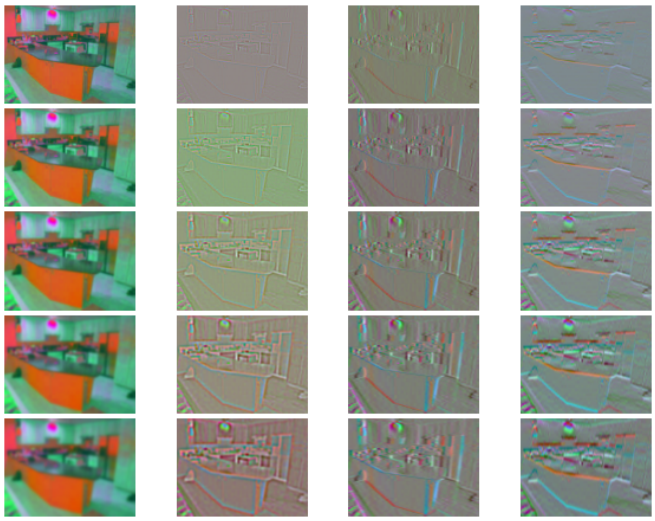
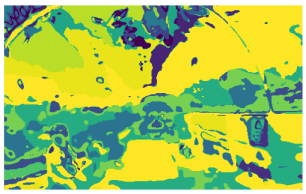


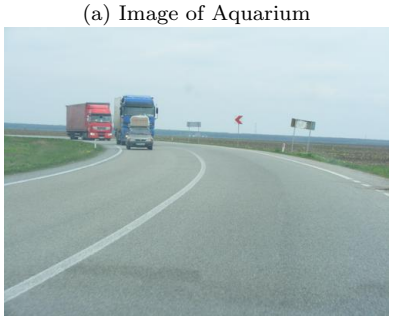
Figure 1: Collage of filter responses with 5 scales

Q1.3

(a) Image of Aquarium



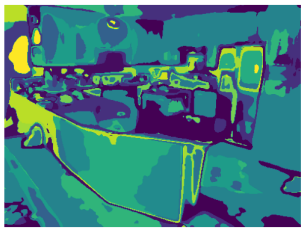
(b) Wordmap of Aquarium



(c) Image of Highway

(d) Wordmap of Highway

(e) Image of Kitchen



(f) Wordmap of Kitchen

The word boundaries for aquarium and kitchen make the most sense to me. Most of these boundaries align with shadows and shades in the image, as well as objects with defined edges. This can as well be said about the highway image. However, the boundary set in the sky doesn't make much sense to me as there isn't a visible difference in terms of pixel value or any physical edge/object present.

Q2.5

Confusion Matrix:

```
[[39.  0.  1.  0.  0.  4.  5.  0.]
 [ 0. 28.  2.  0.  2.  0.  1.  2.]
 [ 0.  2. 32.  1.  1.  1.  1.  8.]
 [ 1. 11.  0. 43. 16.  2.  1.  1.]
 [ 4.  3.  0.  6. 24.  2.  5.  1.]
 [ 1.  1.  5.  0.  2. 37.  7.  5.]
 [ 1.  3.  4.  0.  5.  2. 29.  4.]
 [ 4.  2.  6.  0.  0.  2.  1. 29.]]
 Accuracy: 65.25%
```

Q2.6

Kitchen was the hardest one to classify (as images whose true label was kitchen were mostly predicted as desert and laundromat) as some furniture in kitchen images were white or grey, similar to the washing machines in some laundromat images. This was the same case with desert images as the color of some deserts were similar to the brown drawers/furniture present in some kitchen images.

Q3.1

Starting parameters:
filter-scales = 1 2 3 4 5
K = 20
alpha = 50
L = 2
Accuracy = 55.75%

Changing alpha to 100:
Accuracy = 57.5%
Accuracy increases here because the number of points we retrieve from the filter response of an image increases, thus extracting more features for the dictionary.

Changing K to 30:
Accuracy = 64.5%
Increasing K increases the number of cluster from K-means when grouping the points in filter responses, hence increases the length of the visual words dictionary, thus keeping track of more features.

Changing filter-scales to 1 2 3 4:
Accuracy = 62.5%
Decreasing the number of filter-scales decreases the accuracy because less information is captured at fewer levels.

Changing filter-scales back to 1 2 3 4 5 & alpha to 125 & L to 1:
Accuracy = 64%
Increasing the filter-scale allows us to capture more information at different levels, extracting more features in the filter responses, thus increasing accuracy. Increasing alpha increases the number of points we retrieve from the filter response of an image increases, thus extracting more features for the dictionary, thus increasing accuracy. Since these two measures increases the number of features extracted, increasing L would increase the number of minute features retrieved (since more tiny cells are created in the image to extract features from) that probably contribute to noise which other images can also have, thus decreasing the accuracy. Hence, decreasing L would be prudent.

Changing alpha to 135:
Accuracy = 65.25%
Accuracy increases here because the number of points we retrieve from the filter response of an image increases, thus extracting more features for the dictionary.