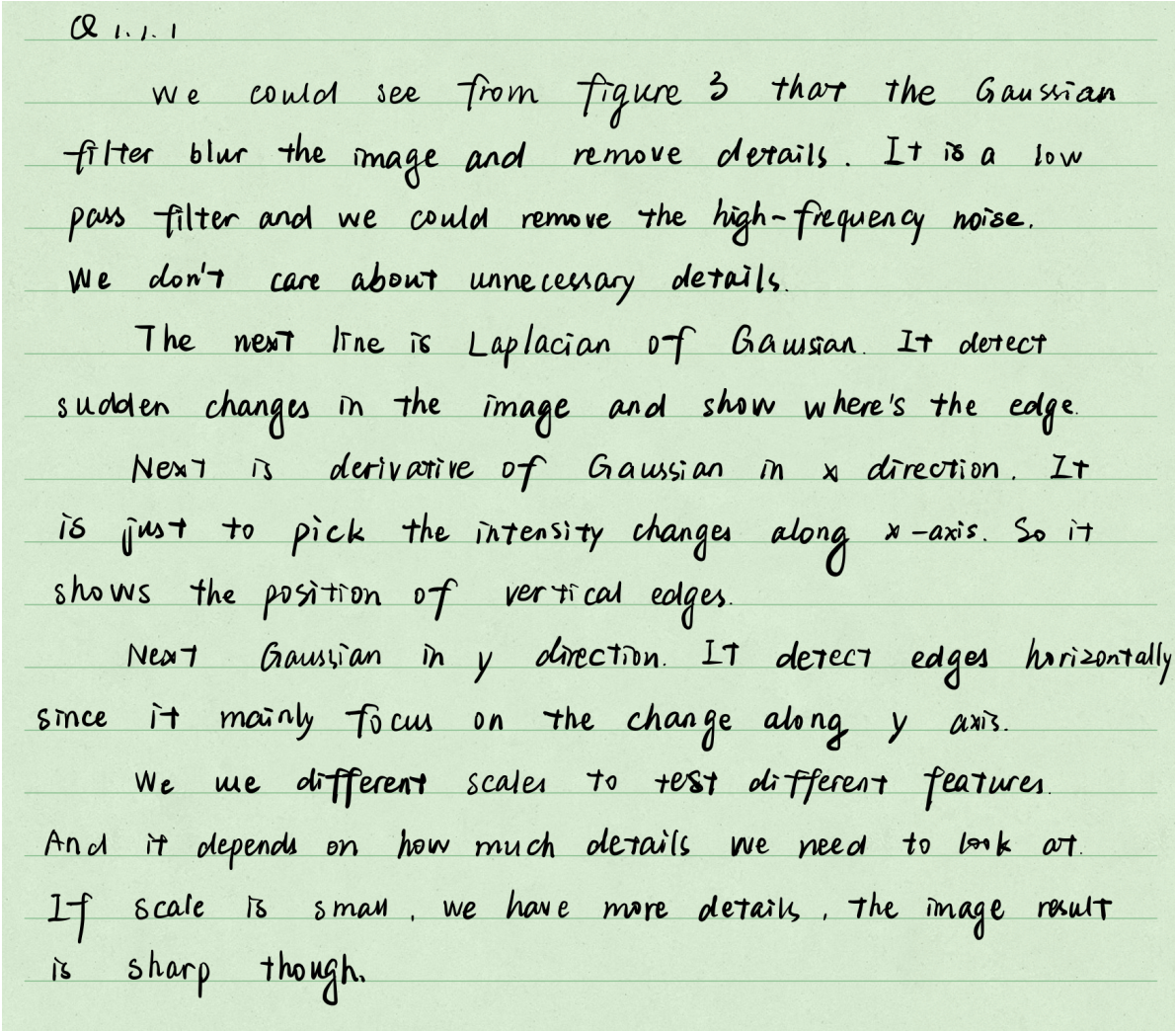
Homework 1 Report

YU YIDUO

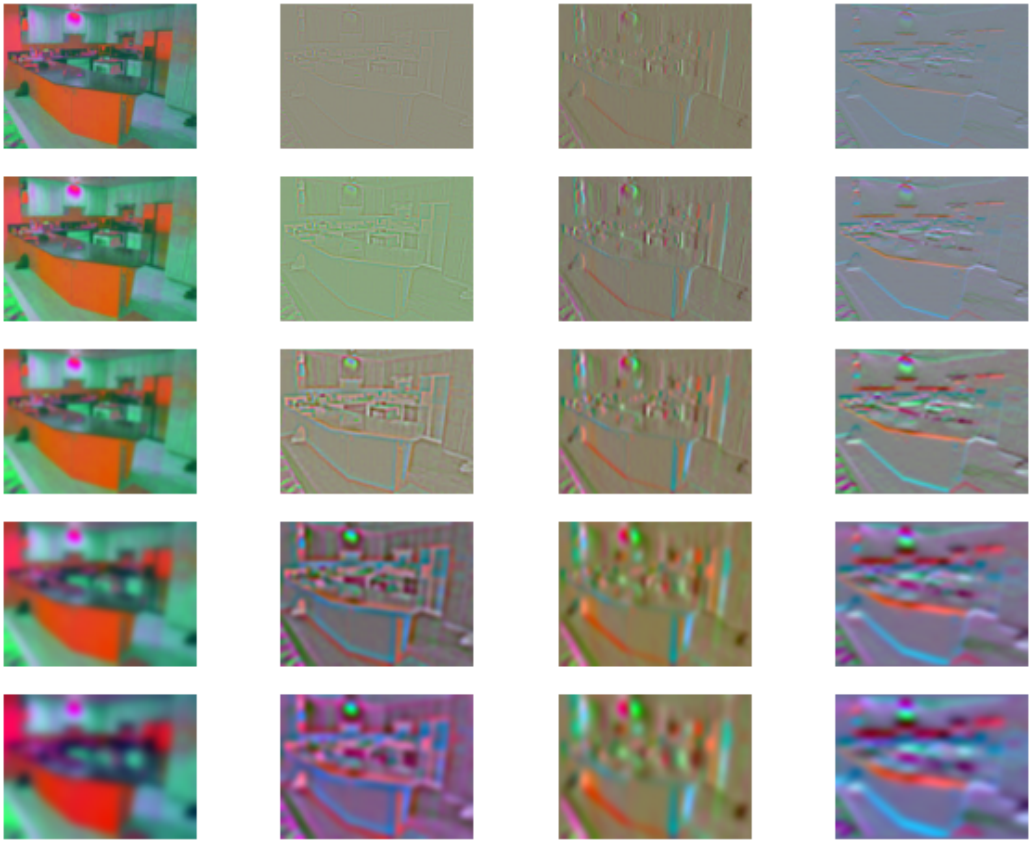
Problem 1.1.1



Problem 1.1.2. Extract filter responses

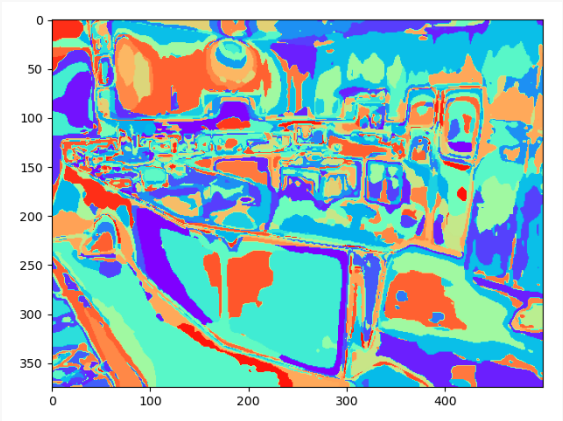
The origin image:

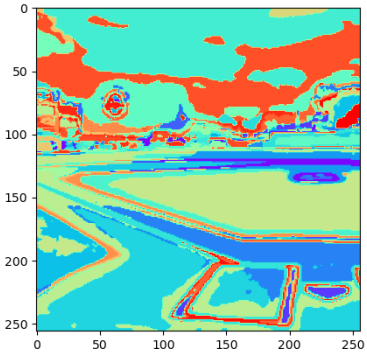
And the extracted filter responses are:

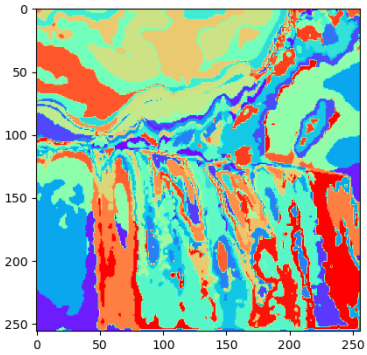


Problem 1.3. Computing Visual Words

We use kitchen, baseball field, and waterfall as the origin images and do the conversion. We could easily see that for the kitchen and waterfall images, there are more features detected. But for the baseball field, less features could be generated. It might because the baseball field has much in common and the color and background are too simple in structure.







Problem 2.5. Quantitative Evaluation

The result confusion matrix is:

```
[[10.  0.  1.  0.  2.  4.  3.  0.]
 [ 1.  9.  1.  1.  0.  0.  1.  7.]
 [ 1.  1.  6.  5.  3.  1.  0.  3.]
 [ 0.  1.  0. 13.  0.  0.  1.  5.]
 [ 7.  1.  0.  0.  9.  2.  1.  0.]
 [ 4.  0.  0.  0.  5.  9.  2.  0.]
 [ 1.  1.  0.  2.  0.  4. 11.  1.]
 [ 0.  1.  1.  5.  0.  0.  2. 11.]]
```

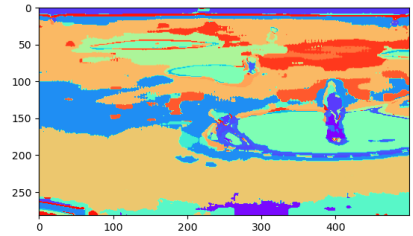
And the result accuracy is:

```
0.4875
```

Problem 2.6. Find the failed cases

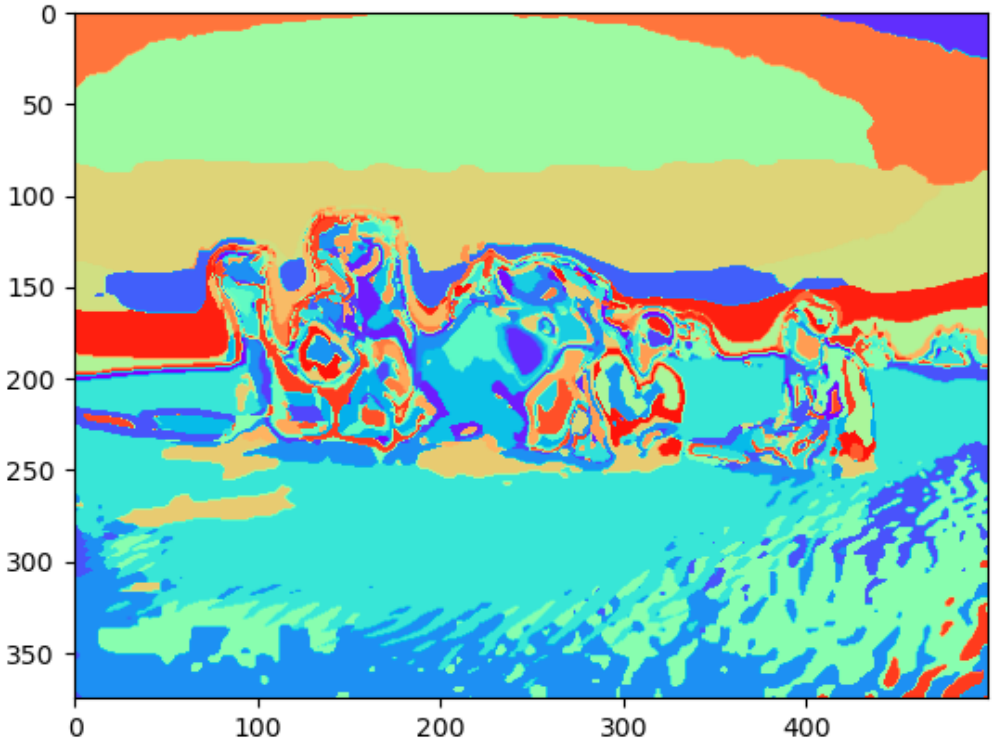
We could see from the confusion matrix that classes in row 2, 3, 5, which are classes “baseball field”, “desert”, "kitchen" are three classes with most prediction errors. By using some technique to save the path of incorrectly predicted samples, we pick three examples below.

First from the second line we see Baseball field has many errors, and most of them are classified as windmill. We could see from the image that because the image is very empty and similar, they have very less features detected. So it is fairly difficult to classify this kind of scenes as the algorithm don't have much to look at.

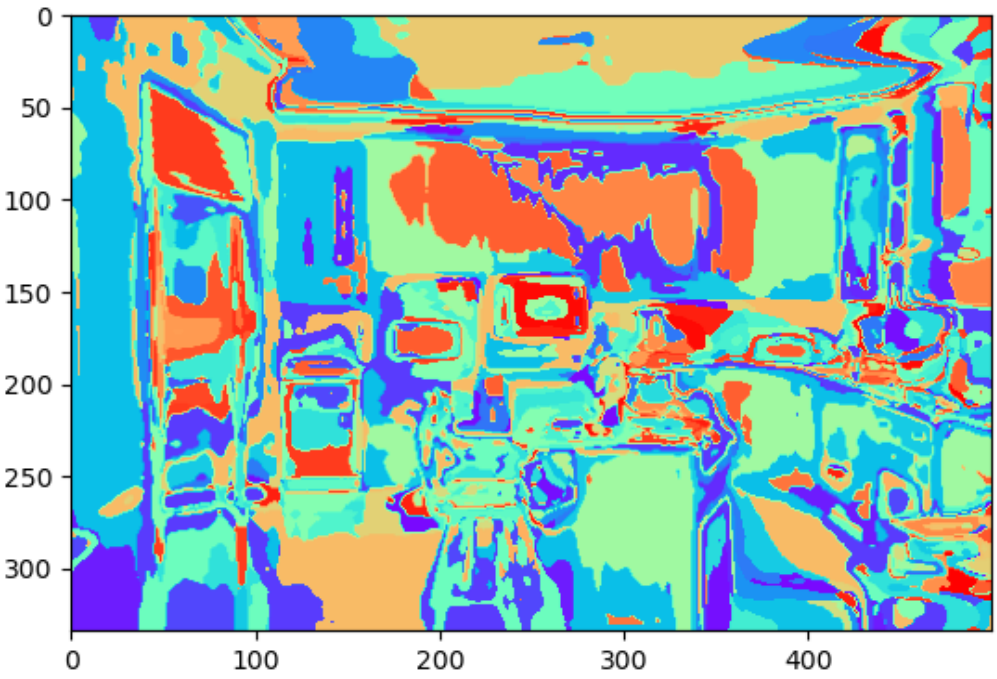


Example of a wrongly predicted baseball field

Desert also has high error. And many of them are predicted as class highway. We could see this is almost the same case with baseball_field since the desert also has very less features to be detected, thus making the classification task very hard.



And then we have kitchen. Why this is misclassified? Because from the given example, we could see that there are many complicated structures in the kitchen. Although there are many features, it is actually a bit similar to the auditorium which also have many structures. So some of our pictures are classified into that class. We could see that the below image is actually very similar to the color and shape of the auditorium.



Problem 3.2. Building a Visual Recognition System: Revisited

The result confusion matrix is:

```
[[19.  0.  0.  0.  1.  0.  0.  0.]
 [ 1. 16.  1.  0.  0.  0.  1.  1.]
 [ 0.  0. 19.  1.  0.  0.  0.  0.]
 [ 0.  0.  0. 20.  0.  0.  0.  0.]
 [ 0.  0.  0.  0. 19.  1.  0.  0.]
 [ 0.  0.  0.  0.  1. 19.  0.  0.]
 [ 0.  0.  1.  0.  0.  0. 19.  0.]
 [ 0.  0.  0.  0.  0.  0.  0. 20.]]
```

And the result accuracy is:

```
0.94375
```

We could easily see that the accuracy is much better in this case. Which is over 90% but the original one is just 45%. This is better since the neural network extracted all the features of the image and the performance is much better than the BOW approach. The CNN and also pooling technique is proved nowadays to be the best tool for computer vision tasks. With the help of Neural network, we could now achieve much higher accuracy for the classification since more features are extracted and represented more accurately.