

16-720 Computer Vision Homework 1

Austin Windham

September 21, 2023

1 Representing the World with Visual Words

1.1 Extracting Filter Responses

Q 1.1.1:

The Gaussian filter smooths the image and emphasizes low frequency features in the image. This filter picks up the overall structure of the image and blurs the fine details.

The Laplace of Gaussian filter shows changes in intensity of the image. This filter picks up on the edges in an image.

The derivative of Gaussian in the x-direction picks up on vertical edges and features in the image.

Alternatively, the derivative of the Gaussian in the y-direction picks up on the horizontal edge and features in the image.

We apply multiple scales of filter responses because different scales can do a better job of picking up certain features for different images. Images are also different sizes and shapes, so certain scales can work better on certain images than others.

Q 1.1.2:

The image below was convolved with the four filters mentioned with filter scale values of 1, 2, 5, and 7.

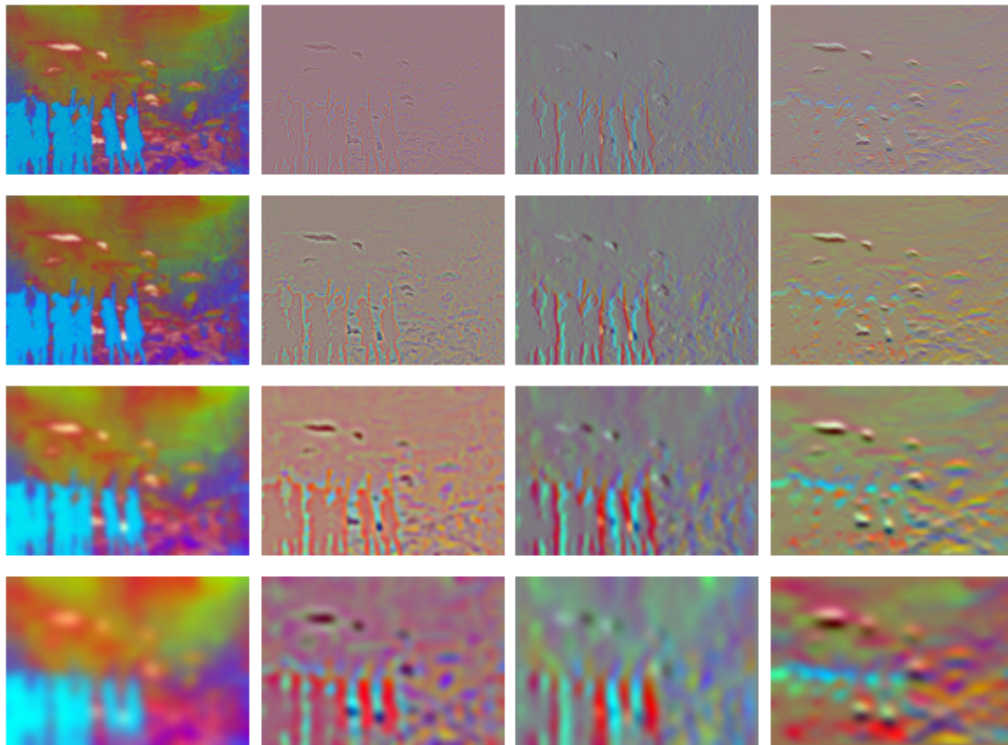


Figure 1: Collage of Convolved Images for aquarium/sun_aztvjgubyrgrup.jpg

1.2 Creating Visual Words

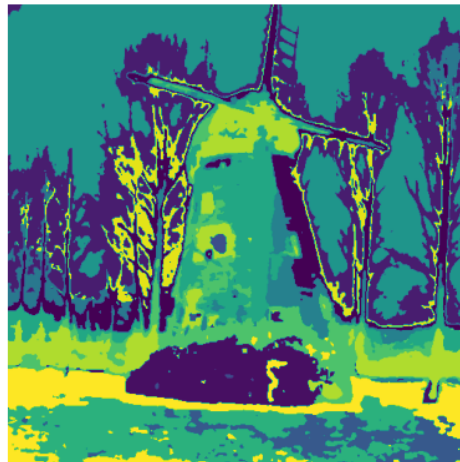
Q 1.2:

Contents are in the code that was submitted.

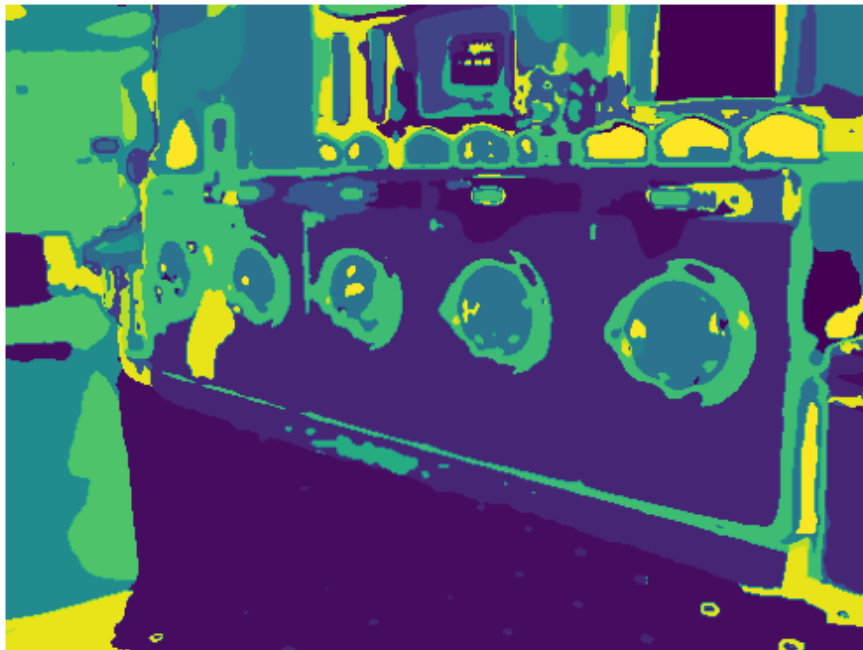
1.3 Computing Visual Words

Q1.3:

Word maps for three images are below. The word maps seem effective and highlight certain features in the images. They seem to show contours and edges as well as changes in color and intensity.



Figures 2 and 3: Original and Wordmap image of windmill/sun_agudhwulyxcizdjv.jpg



Figures 4 and 5: Original and Wordmap Image of laundromat/sun_aabvooxzwmzzvwds.jpg



Figures 6 and 7: Original and Wordmap Image for highway/sun_aacqsbumiuidokeh.jpg

2 Building a Recognition System

2.1 Extracting Features

Q2.1:

Contents are in the code that was submitted.

2.2 Multi-Resolution: Spatial Pyramid Matching

Q 2.2:

Contents are in the code that was submitted.

2.3 Comparing Images

Q2.3:

Contents are in the code that was submitted.

2.4 Building a Map of the Visual World

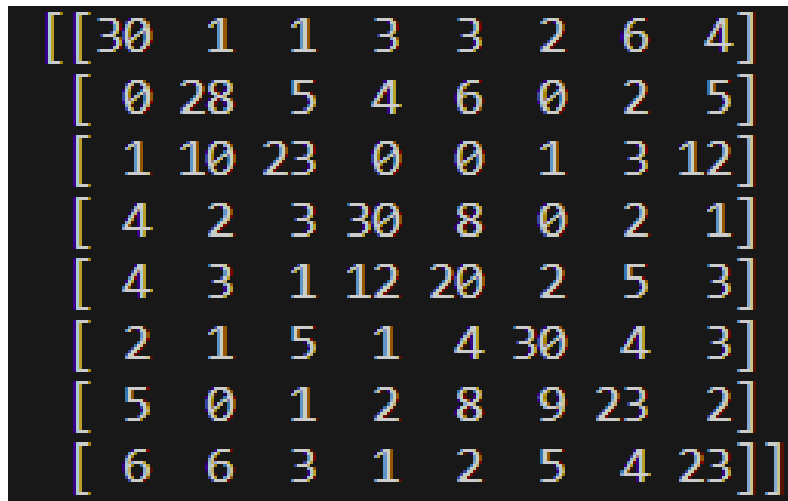
Q2.4:

Contents are in the code that was submitted.

2.5 Quantitative Evaluation

Q2.5:

With the default parameters of filter scales = [1, 2], K = 10, L = 1, and alpha = 25, my confusion matrix is given below.



A confusion matrix displayed on a black background with yellow and orange numbers. The matrix is an 8x8 grid enclosed in large square brackets. The values are as follows:

30	1	1	3	3	2	6	4
0	28	5	4	6	0	2	5
1	10	23	0	0	1	3	12
4	2	3	30	8	0	2	1
4	3	1	12	20	2	5	3
2	1	5	1	4	30	4	3
5	0	1	2	8	9	23	2
6	6	3	1	2	5	4	23

Figure 8: Confusion Matrix for Default Parameters

Accuracy = 51.75 %

2.6 Find the Failures

Q2.6:

Based on the confusion matrix, a significant number of laundromats were incorrectly classified as kitchens. This is probably a result of the two classifications having similar looking features from having appliances that are similar looking gray metal boxes like an oven, dishwasher, dryer, or washing machine. They both may also have similar features when some of the filters are applied since the two scenes both have a lot of horizontal edges like countertops, table tops, and large appliances. These features can be seen in the images below of kitchen and laundromat files.



Figures 9 and 10: kitchen/sun_abjllhhfuvcygwtk.jpg and laundromat/sun_aabvooxzwmzzvwds.jpg

Another scene that is harder to classify are highways, and the misclassifications of highways are often windmills. From some of the images, I believe this is most likely due to the fact that some of the highways are surrounded by grass and trees with a large view of the sky and some structures in the middle, which is similar to the windmill scenes where they are

surrounded by trees and grass and have usually a grayer vertical structure in the middle and sky in the background. This is emphasized by the figures below.



Figures 11 and 12: windmill/sun_agudhwulyxcizdjv.jpg and highway/sun_advstbacygihnsur.jpg

3 Improving Performance

3.1 Hyperparameter Tuning

Q3.1:

The table below outlines my steps for parameter tuning. I started with the default parameters and an accuracy of 51.75 % and then increased L to 3 yielding an accuracy of 52.75 %. I then increased K to 20, and the model's accuracy increased to 59.00 %. I then changed the filter scales to [1, 2, 5, 10], and the accuracy increased again to 60.75 %. After that, I increased alpha to 40, but the accuracy decreased to 58.00 %, so I returned alpha back to 25 and increased K to 30, yielding an accuracy of 62.75 %. I increased K again to 50, but the accuracy decreased to 61.50 %, so my final maximum accuracy was 62.75 % with K at 30.

Increasing K appeared to be the most helpful, which makes sense because an increased K equates to more dictionary words that can be used to distinguish pictures. Once the K became too large, there was a decrease in accuracy because K was probably making more words than were necessary and breaking up well-defined clusters to generate the additional clusters to reach K-numbered clusters. Applying more filter scales also helped the model because more filter scales picks up on more features that can be used to distinguish the images. Increasing the number of layers, L, also increased the accuracy slightly, which was expected because a greater L equates to more information about the spatial structure of the image. I thought that an increased alpha would increase the accuracy since the system would be sampling more pixels, but it had the opposite effect. This could mean that the default alpha was already great enough or that the bag of words approach is not affected much by an increased alpha.

Parameter Tuning Ablation Table

Step	L	K	Filter Scales	Alpha	Accuracy
1	1	10	[1, 2]	25	51.75 %
2	3	10	[1, 2]	25	52.75 %
3	3	20	[1, 2]	25	59.00 %
4	3	20	[1, 2, 5, 10]	25	60.75 %
5	3	20	[1, 2, 5, 10]	40	58.00 %
6	3	30	[1, 2, 5, 10]	25	62.75 %
7	3	50	[1, 2, 5, 10]	25	61.50 %

Maximum Accuracy = 62.75 %