CSE 573

Spatial Pyramid Matching for Scene Classification

Project Report

Pranav Jain

Person Number : 50208349

UBIT : pjain4

CSE 573

Spatial Pyramid Matching for Scene Classification

Project Report

Introduction

The purpose of this project was to implement a scene classification system that uses the bag-of-words approach with a spatial pyramid extension. This document is a write up of the project and contains implementation details, results and their analysis.

Representing the World with Visual Words

Q1.0 What properties do each of the filter functions pick up?

There are 4 distinct filters that we are using for obtaining features and these features are taken at 5 different scales (depending on the σ value) with different values of σ from the list [1,2,4,8,11.31].

The types of filters that we use are given below

- Gaussian: By looking at Figure 1 Top Left and bottom first row we can see that this filter blurs the image. As the scale increases so does the blurriness. By looking at the top right part in the image we can see that it is like a weighted average which weighs the central values more heavily.

- Laplacian of Gaussian: As can be seen from taking a look at the image and second row of the montage, the filtered images outline the edges in the image and by looking at the filter we can see that response will be almost 0 in places where the intensities are uniform. As we increase σ the edges smoothen and some of them are lost.

- δx: This filter brings out vertical edges in the images. This can be clearly seen from the top right image as we will only get a high response when there is change in intensities horizontally which gives a vertical edge.

- δy: This filter brings out horizontal edges in the images. This can be clearly seen from the top right image as we will only get a high response when there is change in intensities vertically which gives a horizontal edge.

Q1.1 Submit the collage of 20 images in the write-up

The result is included as a montage in the bottom part of Figure 1 on page 10 .The first row shows the results with a Gaussian filter, the second shows results with Laplacian of Gaussian and the third and fourth are with δx and δy respectively.

Q1.3 Map each pixel in the image to its closest word in the dictionary and Visualize three wordmaps of three images from any one of the category

The images have been selected from the garden category and their wordMap has been visualized using imagesc function of matlab in Figure 2 on page 11. Since the dictionary that is used has 150 words , words from 1-150 are all associated with different colors from the jet colrmap which can be seen in the extreme right of the figure. This sort of a representation helps us visualize that pixels that have same color got matched to the the same word from the dictionary and might be similar features.

Extracting Features

Q2.5 Quantitative Evaluation

The Dictionary was made by extracting α (#100) 60 length features from each training image(#1349) and then doing a K-means clustering on these selected features to obtain K words (#150). The system was then tested on 160 test images and the confusion matrix is given in Table 1 on page 6.

The accuracy obtained with the following system is 47.5%. Without spatial pyramid matching the accuracy dropped to 46.25%. Mapping from labels to categories is given in table 4.

Q2.6 List some of the classes/samples and discuss why they are more difficult to classify than the rest.

Some cases are more difficult than others, for example images belonging to the category 'art gallery' have been wrongly classified the most with an accuracy of only 15%. Some wrongly classified art galleries and their matches are showed in figure 3 on page 12. In the first row they presence of chairs could have been the reason, In row 2 it may be due to similar floors and in rows 3 and 4 it may be due to similarities in color (grey and white) respectively.

It is also worth noting that oceans were classified with an accuracy of 75% and this may be due to presence of mainly one color and on the other hand 30% of tennis courts were classified as oceans and only 20% of tennis courts were rightly identified. This again may be chalked up to presence of mainly one color/ images being texture less.

Q2.7 Improving Performance

Code for this section is in the custom folder. The custom folder contains two folders, custom225 and custom300 which have dictionary sizes of 225 and 300 respectively.

Parameter Tweaking.

- Spatial Pyramid Matching: This section presents results obtained with different values of L for spatial pyramid matching ranging all the way from [0-4], where 0 is equivalent to not having spatial pyramid matching. The results are presented in figure 4 on 13. The Highest accuracy was 49.38% seen when L=3 and on further increasing the value of L there was a drop in accuracy to 43.75% for K=150 and α = 100. This was also observed for dictionary sizes of 225 which got the highest accuracy of 52.50% at L=3 and fell to 47.50% at L=4. The dictionary size of 300 got it's maximum accuracy at L=2 at 54.37% and showed the same trend of falling at higher values of L. The reason for a decrease in accuracy at higher values could be because the highest level of the pyramid is too finely subdivided.

- Dictionary Size: This sections looks at how dictionary sizes effect the accuracy. From figure 4 we can see that the maximum accuracy was obtained for the highest dictionary size (300) at L=2 (54.37%). While for dictionary sizes 150 and 225 the accuracy was 47.5% and 50.62% respectively. It appears that a larger dictionary size gives a better accuracy for most cases. The confusion matrix for the highest accuracy is given in table 3 on 8

Actual | Predicted	1	2	3	4	5	6	7	8
1	3	4	0	2	7	1	2	1
2	1	8	2	2	4	0	1	2
3	0	0	14	0	1	1	3	1
4	2	2	1	12	3	0	0	0
5	2	4	0	2	11	0	1	0
6	0	0	4	0	2	9	5	0
7	0	0	0	0	0	4	15	1
8	2	2	3	1	1	1	6	4

Table 1

Confusion Matrix: C(i,j) gives the number of images of category I that got identified as that of category J. Dictionary size = 150,α=100,L=2

1	2	3	4	5	6	7	8
art gallery	computer room	garden	ice skating	library	mountain	ocean	tennis court

Table 2

The Mapping from category to Labels

Actual | Predicted	1	2	3	4	5	6	7	8
1	5	1	0	6	5	0	1	2
2	0	8	1	2	7	1	1	0
3	0	1	15	2	1	0	1	0
4	0	0	1	17	2	0	0	0
5	5	4	1	1	9	0	0	0
6	0	0	2	0	1	9	8	0
7	0	0	0	0	0	2	18	0
8	1	3	2	2	2	1	3	6

Table 3

Confusion Matrix: C(i,j) gives the number of images of category I that got identified as that of category J. Dictionary size = 300,α=100,L=2

1	2	3	4	5	6	7	8
art gallery	computer room	garden	ice skating	library	mountain	ocean	tennis court

Table 4

The Mapping from category to Labels

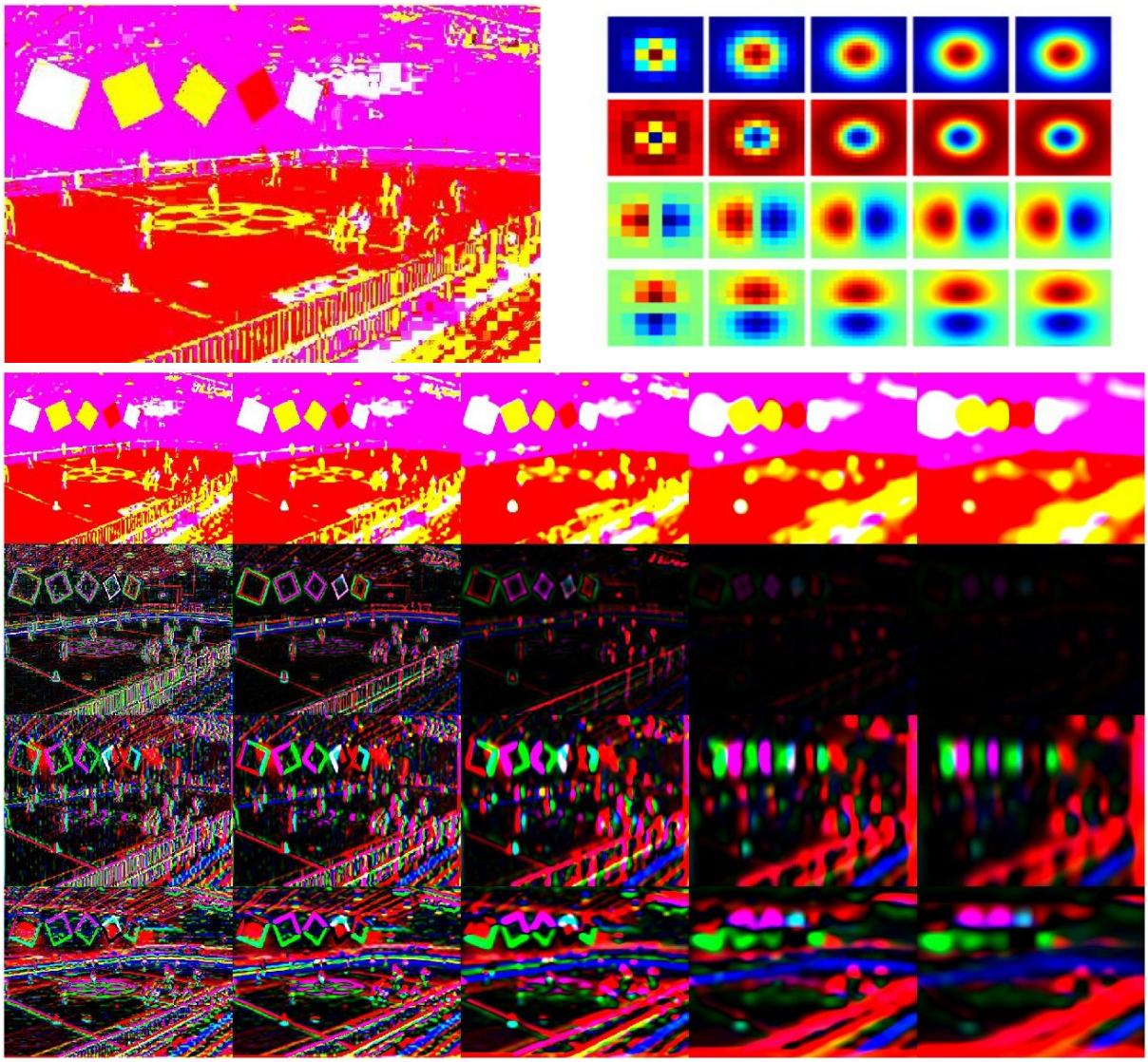

Figure 1. Top Left: Image in L,a,b space. Top Right:Filter Bank Bottom: Montage of image after convoluted with the filters.

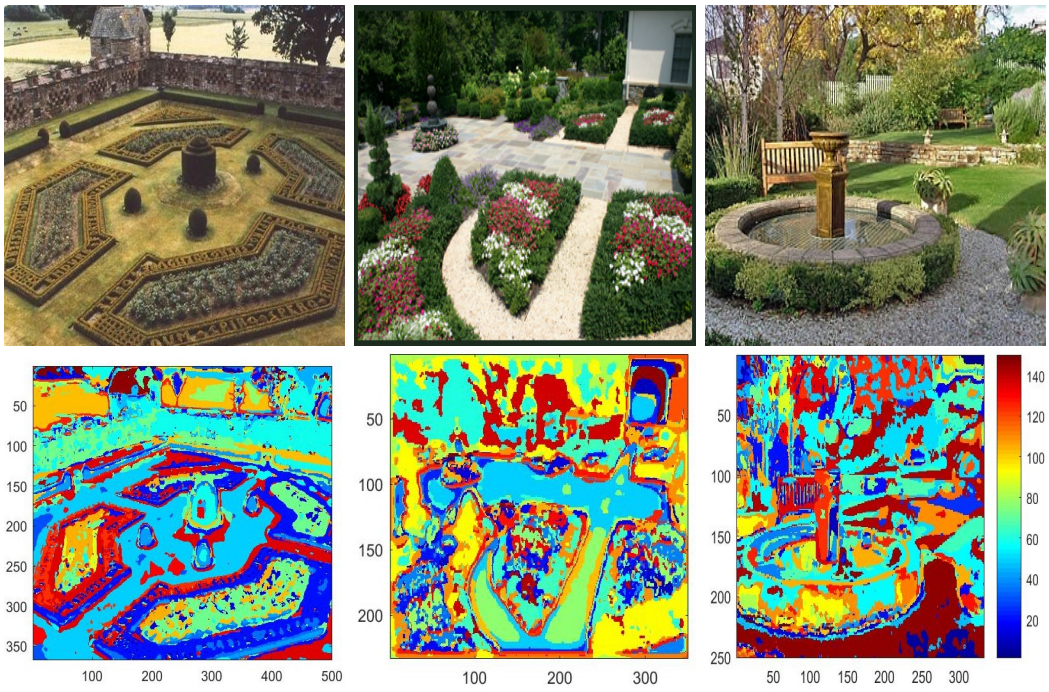

Figure 2. Top Row: Images selected from the garden category. Bottom Row: Wordmap Representation of these images. Extreme Right: Color map

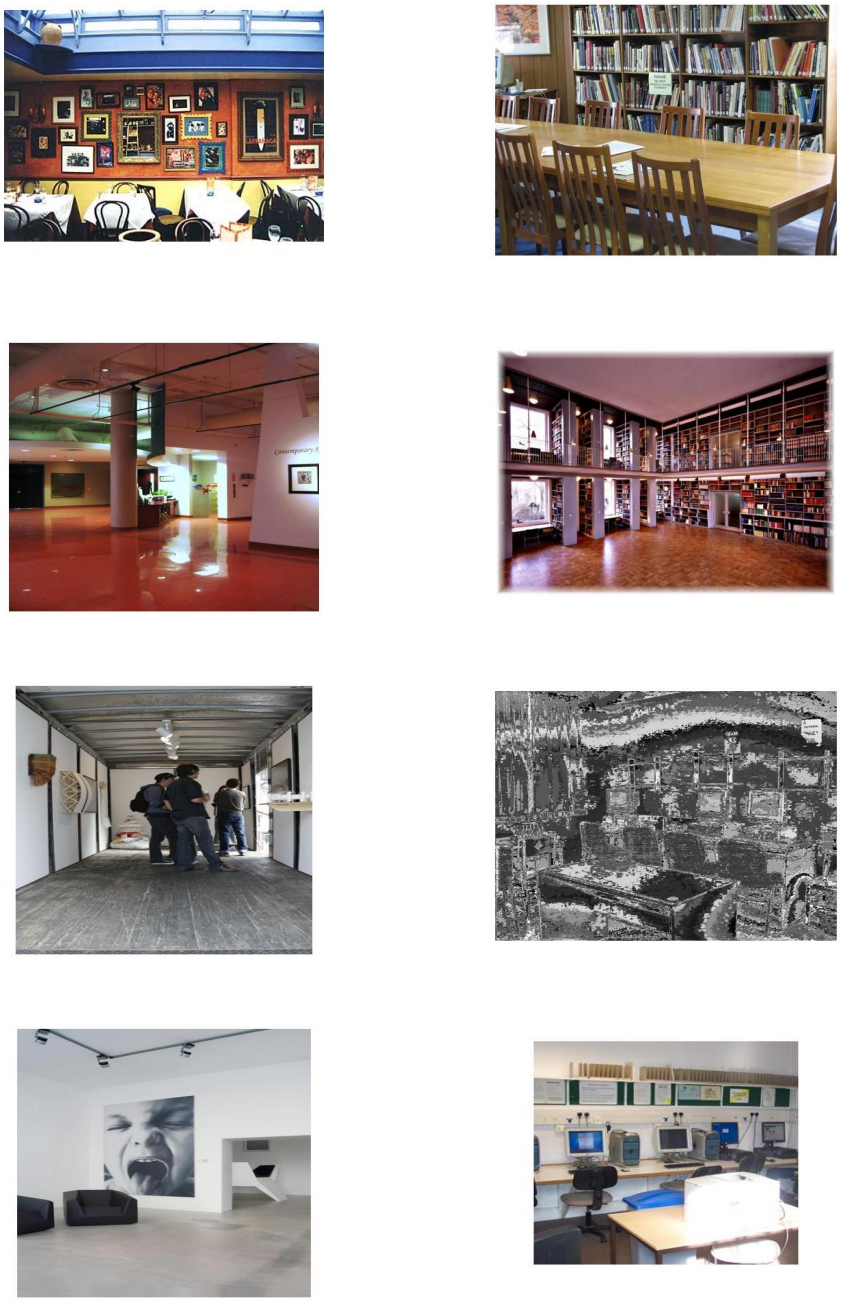

Figure 3. Left Column: Images of Art Galleries. Right Column: Corresponding Images that they were found closest to

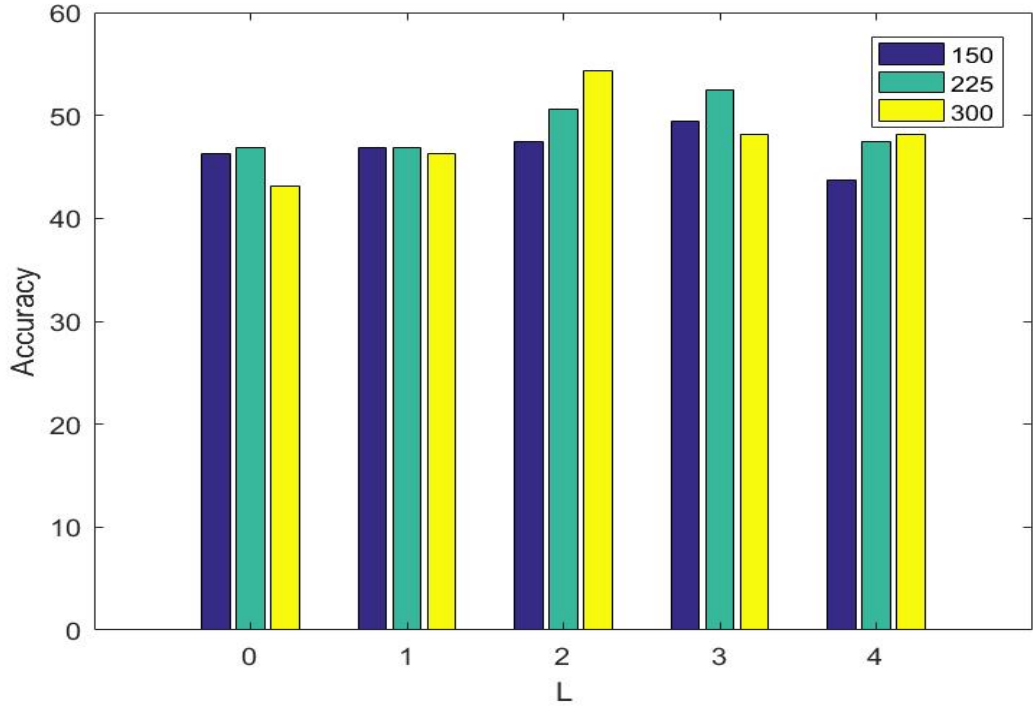

Figure 4. Accuracy of the System vs L for different dictionary sizes.