DLCV hw2 b04901155 鄭閎

Problem 1

$$K(x, x') = \Phi(x)^T \Phi(x') = (x^T x')^2$$

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad x' = \begin{bmatrix} x_1' \\ x_2' \end{bmatrix} \quad K(x, x') = (x_1 x_1' + x_2 x_2')^2 = x_1^2 x_1'^2 + 2 x_1 x_1' x_2 x_2' + x_2^2 x_2'^2 = \Phi\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)^T \Phi\left(\begin{bmatrix} x_1' \\ x_2' \end{bmatrix}\right)$$

$$x_1^2 x_1'^2 + 2 x_1 x_1' x_2 x_2' + x_2^2 x_2'^2 = \begin{bmatrix} x_1^2 & \sqrt{2} x_1 x_2 & x_2^2 \end{bmatrix} \begin{bmatrix} x_1'^2 \\ \sqrt{2} x_1' x_2' \\ x_2'^2 \end{bmatrix} = \Phi\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)^T \Phi\left(\begin{bmatrix} x_1' \\ x_2' \end{bmatrix}\right)$$

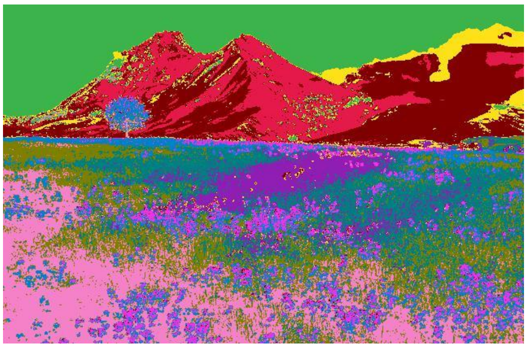
$$\Rightarrow \Phi(x) = \Phi\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1^2 \\ \sqrt{2} x_1 x_2 \\ x_2^2 \end{bmatrix}$$

Problem 2

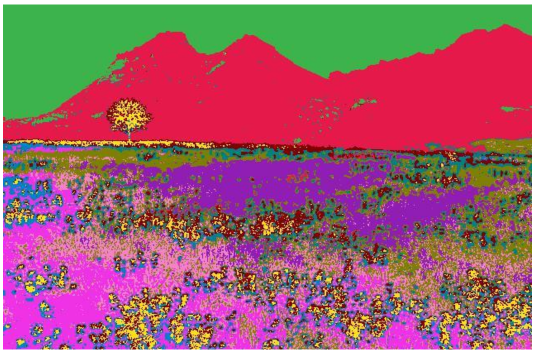
Original image

(a) Color segmentation

(i) RGB

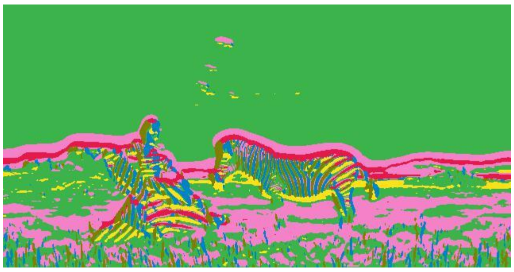


(ii) Lab color space

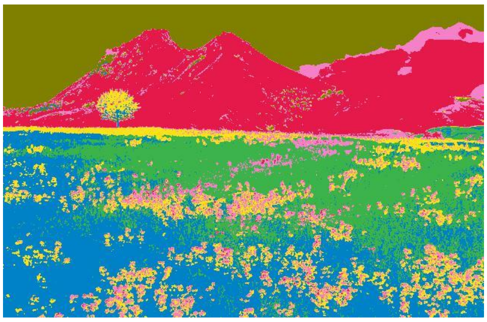
	

(b) Texture segmentation

(i) texture(38 dims)

(ii) combine texture and color (42 dims)

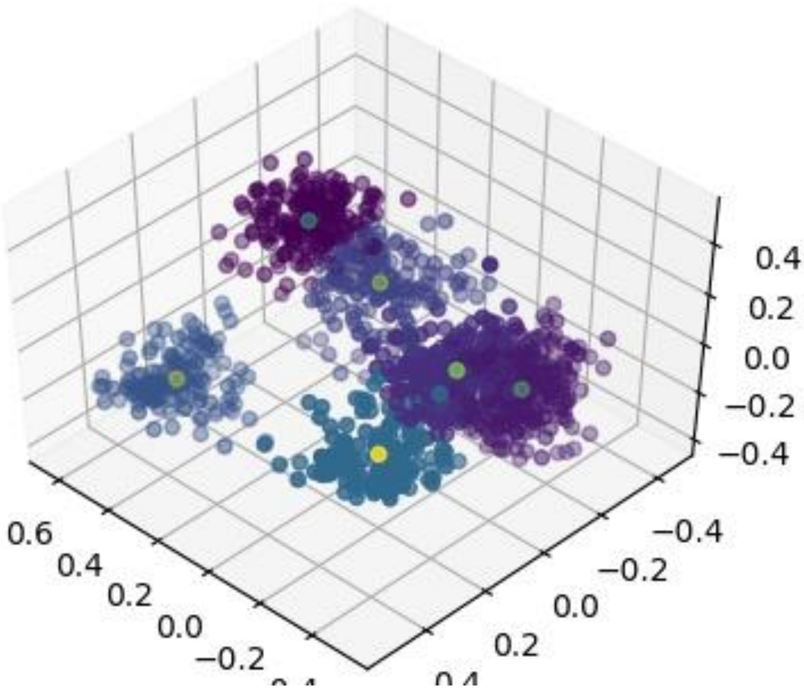
	

Problem 3

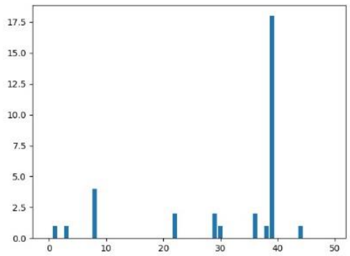
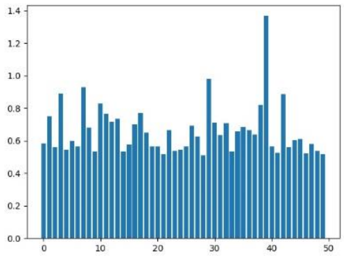
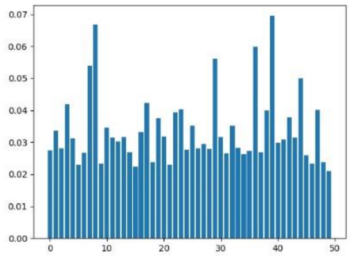
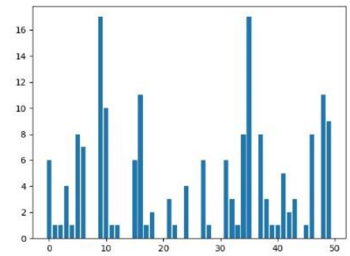
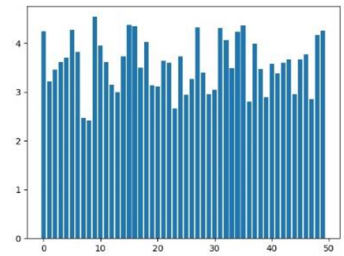
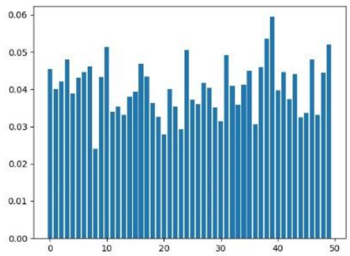
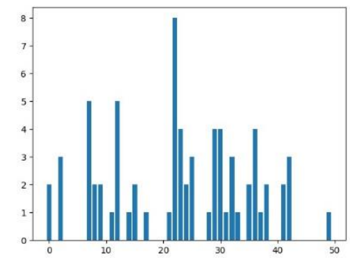
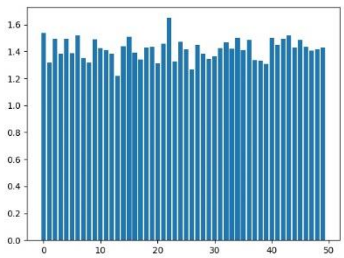
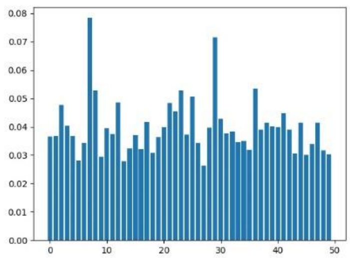
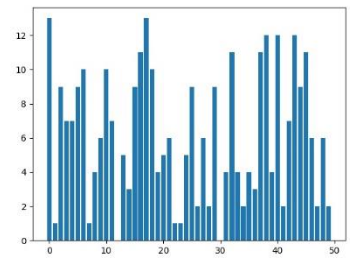
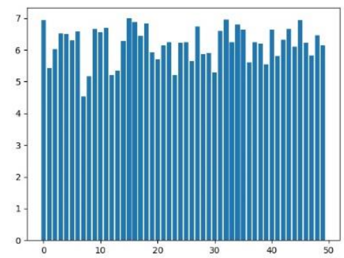
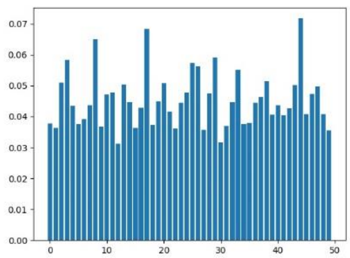
(a) Image picked: Suburb/image_0029.jpg

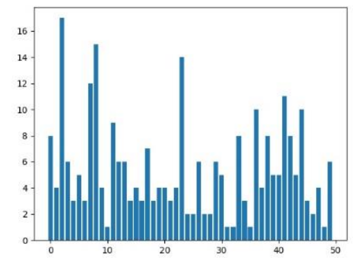
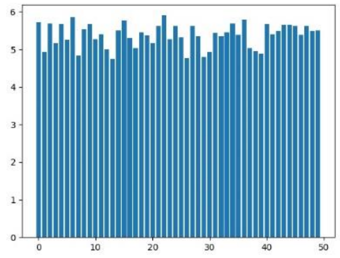
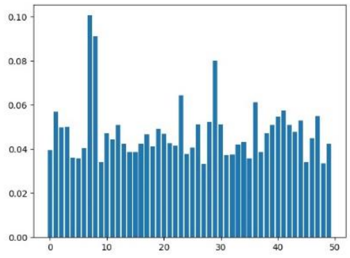
Hessian Threshold:6000   30 interest points detected

(b) Different colors are used to distinguish different clusters. In each cluster there is a point to represent the centroid.



(c) Bar chart of different BoW strategy over 5 different categories of image
Image picked: Coast/image_0032.jpg, Forest/image_0003.jpg, Highway/0010.jpg, Mountain/image_0044.jpg, Suburb/image_0029.jpg

<table>
<tr><td></td><td>hard sum</td><td>soft sum</td><td>soft max</td></tr>
<tr><td>Coast</td><td>
</td><td>
</td><td>
</td></tr>
<tr><td>Forest</td><td>
</td><td>
</td><td>
</td></tr>
<tr><td>Highway</td><td>
</td><td>
</td><td>
</td></tr>
<tr><td>Mountain</td><td>
</td><td>
</td><td>
</td></tr>
</table>

Suburb			

I would expect the hard sum strategy to have better classification result. From the plot we can observed that the distribution of the result from hard sum differs a lot among different classes. Therefore, the classification would be easier due to the stronger characteristics.

(d) Accuracy

(i) Train-10, # of cluster = 50, max iteration = 5000

BoW strategy	hard sum	soft sum	soft max
accuracy	0.564	0.356	0.576

The accuracy of hard sum is very close to that of soft max. I think the reason that soft max can have good performance is that it reflects the different importance of the interest point in each visual word. Unlike hard sum that regards all the chosen interest point as same importance. Therefore, soft max keeps more flexibility.

The reason that soft sum fails is that the distribution cannot show strong difference among different class, thus it’s hard to be used for classification.

(ii) Train-100, # of cluster = 50, max iteration = 5000

BoW strategy	hard sum	soft sum	soft max
accuracy	0.69	0.63	0.698

# of cluster = 100, max iteration = 5000

BoW strategy	hard sum	soft sum	soft max
accuracy	0.678	0.636	0.704

# of cluster = 200, max iteration = 5000

BoW strategy	hard sum	soft sum	soft max
accuracy	0.554	0.64	0.738

# of cluster = 200, max iteration = 10000

BoW strategy	hard sum	soft sum	soft max
accuracy	0.586	0.64	0.708

# of cluster = 300, max iteration = 10000

BoW strategy	hard sum	soft sum	soft max
accuracy	0.476	0.636	0.738

Using Train-100 shows improved results over Train-10 on all 3 BoW strategies. This is quite intuitive that more train data results in a better outcome.

I also change the number of cluster and max iteration of k-means. From the result I observed that as number of cluster increase, the performance of soft max is better while hard sum is worse. A simple explanation is that hard sum get overly “distracted” by too many visual words. On the other hand, soft max can get more flexibility with the number of visual words increase and thus create a better model.