CommE 5052: Deep Learning for Computer Vision **Spring 2018**

Homework 2

Name: Zhixiang Wang *Student ID: R06944046*

Note: This solution is based on `Matlab`, using the libs/functions including `extractFeatures`, `pca`, `kmeans` and `fitcknn`, etc.

Problem 1: Kernel Trick

Given $\mathbf{x} = [x_1 \ x_2]^T$ and

$$K(\mathbf{x}, \mathbf{x}') = \Phi(\mathbf{x})^T \Phi(\mathbf{x}')^T = \left(\mathbf{x}^T \mathbf{x}'\right)^2 \tag{1}$$

we can derive that

$$K(\mathbf{x}, \mathbf{x}') = {x_1}^2 {x_1'}^2 + {x_2}^2 {x_2'}^2 + 2x_1 x_1' x_2 x_2' = \begin{bmatrix} x_1^2 & x_2^2 & \sqrt{2}x_1 \, x_2 \end{bmatrix} \begin{bmatrix} {x_1'}^2 & {x_2'}^2 & \sqrt{2}x_1' \, x_2' \end{bmatrix}^T \tag{2}$$

hence, the explicit form of the mapping $\Phi(\mathbf{x}) : \mathbb{R}^2 \to \mathbb{R}^3$ is

$$\Phi(\mathbf{x}) = \begin{bmatrix} x_1^2 & x_2^2 & \sqrt{2}x_1 \, x_2 \end{bmatrix}^T \tag{3}$$

Problem 2: Color and Texture Segmentation

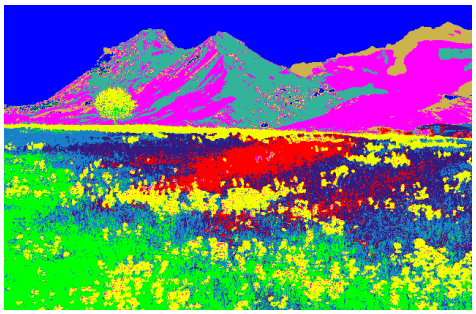
(a) **Color segmentation: i)** The color segmentation results for both images (in RGB space) are shown in Figure 1(a-b). **ii)** The color segmentation results for both images (in Lab space) are shown in Figure 1(c-d).

(b) **Texture segmentation:** The texture segmentation results for both images are shown in Figure 1(e-f). **ii)** The mixed segmentation results for both images are shown shown in Figure 1(c-d). We combine both color (Lab color space, using `rgb2lab`) and texture features together.

Problem 3: Recognition with Bag of Visual Words

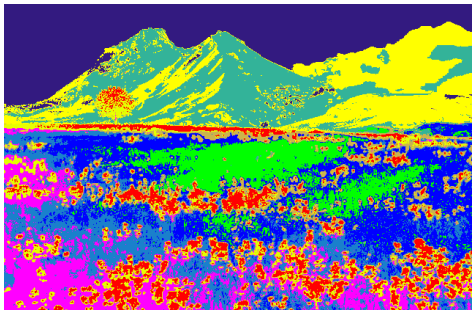
(a) Image, which is randomly picked, with the 30 most dominant interest points detected using SURF is shown in Figure 2(a).

(b) We randomly select 6 clusters and plot the visual words (marked with '+') and the associated interest points (marked with '·') in this 3D PCA subspace, as shown in Figure 2(b). As required, the same color is used to denote projected visual words and interest points and different colors represent different clusters.



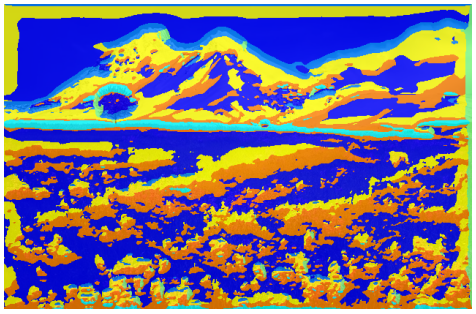
(a) RGB color segmentation

(b) RGB color segmentation

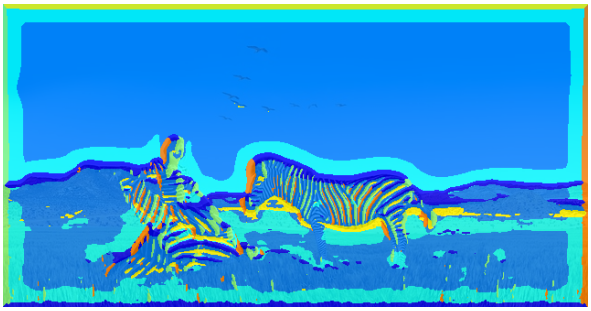


(c) Lab color segmentation

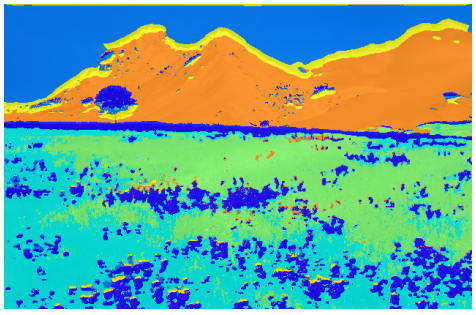
(d) Lab color segmentation



(e) Texture segmentation



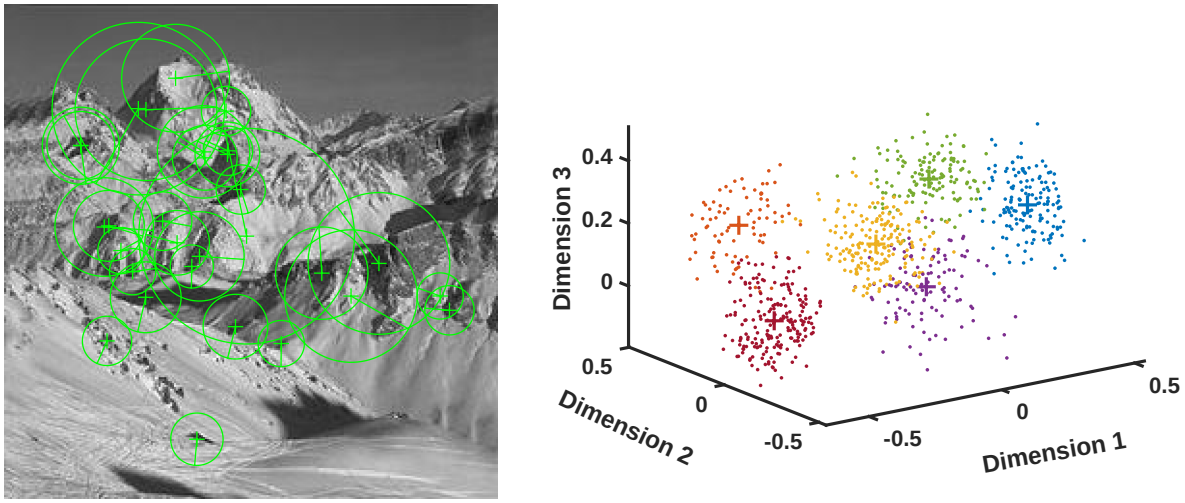
(f) Texture segmentation



(g) Color+texture segmentation

(h) Color+texture segmentation

Figure 1: Results of **Problem 2**.



(a) (b)

Figure 2: Results of **Problem 3** (a) and (b).

(c) The results are shown in Figure 3, it looks like that **Soft-max** is better. Because its frequency has almost two values, therefore, we can determine its category from these features. But **Soft-sum**, its frequency has almost one values and **Hard-sum** its frequency has multiple values, which is hard for us to conduct classification.

(d) The the classification accuracy is shown in Table 1. **i)** The results (first row) are as expected: **Soft-max** is better for classification. **ii)** We can observe the improvement in the accuracy. We try several different parameters pairs (# of visual words, # of iteration, # of k nearest neighbors).

Table 1: Results of **Problem 3** (d). Parameters pair in the bracket is (# of visual words, # of iteration, # of k nearest neighbors).

	Hard-Sum	**Soft-Sum**	**Soft-Max**
Train-10	0.520	0.518	**0.592**
Train-100 $(50, 5000, 5)$	0.582	0.656	**0.682**
Train-100 $(250, 5000, 5)$	0.512	0.652	**0.728**
Train-100 $(500, 5000, 5)$	0.512	0.656	**0.756**
Train-100 $(500, 5000, 10)$	0.482	0.650	**0.758**
Train-100 $(500, 5000, 15)$	0.482	0.624	**0.770**
Train-100 $(500, 5000, 20)$	0.452	0.620	**0.766**

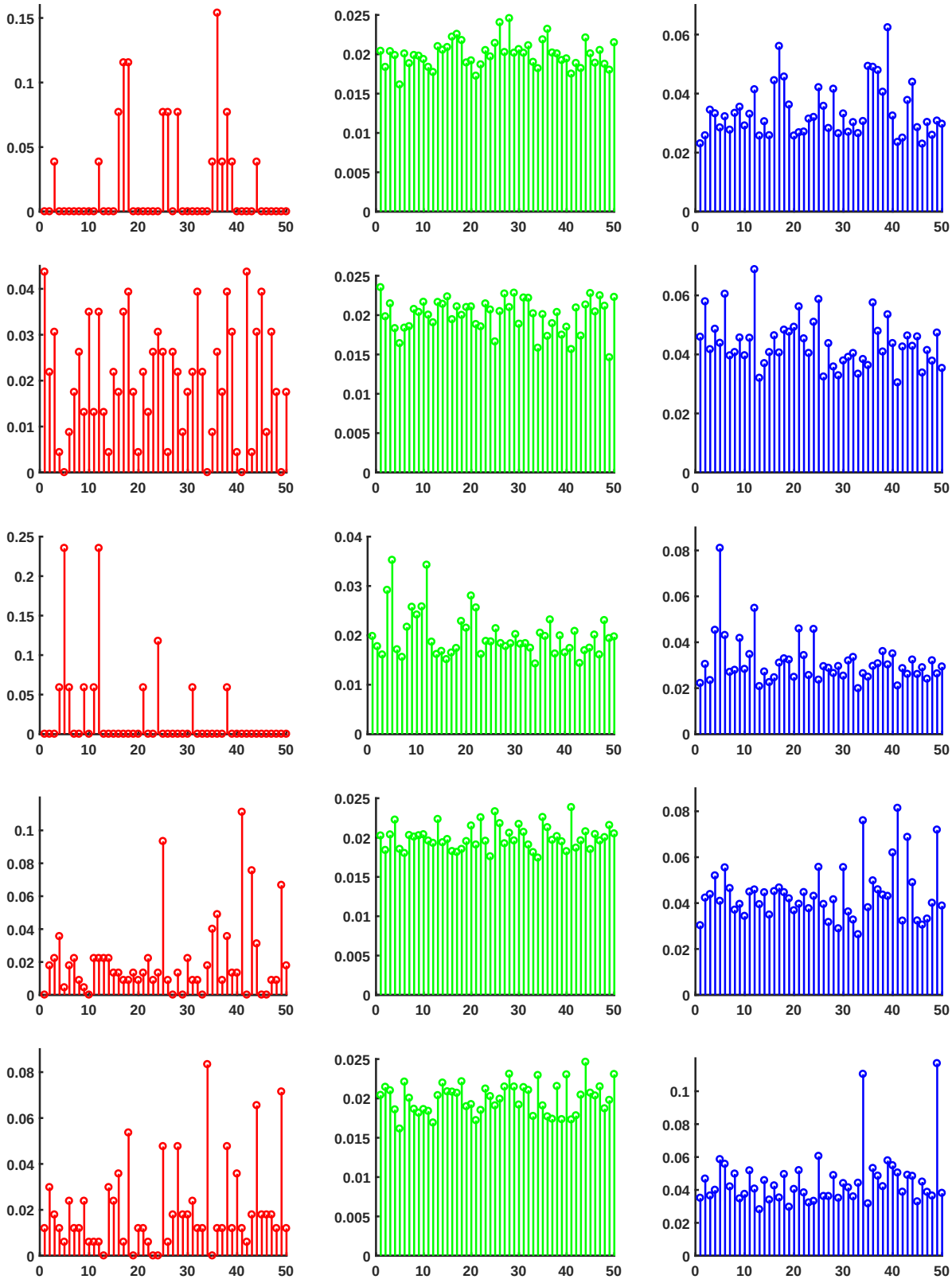


Figure 3: Results of **Problem 3** (c). (left) Hard sum, (middle) Soft sum, (right) Soft max. From top to bottom, Coast, Forest, Highway, Mountain, and Suburb respectively.