浙江大学实验报告

课程名称：专题研讨　　指导老师：Ping TAN

实验项目名称：Project 4: Scene recognition with bag of words

学生姓名：顾继庠　专业：计算机与科学技术　学号：3150105385

实验日期：2019 年 6 月 26 日

1. Experiment Task and Result (Brief description)

Q1.1 Extract Filter Responses

Original Image

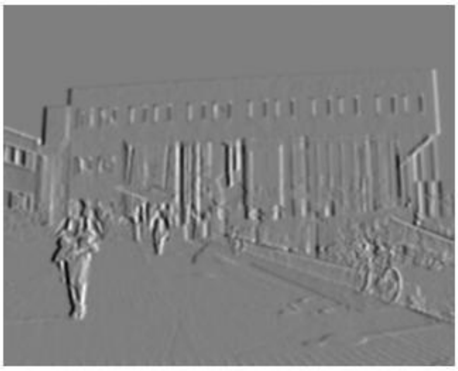
(1) Gaussian

Gaussian: Smoothed edges, loss of high frequency detail in the image (it seems detail is less granular post Gaussian transform)

(2) Laplacian:

Laplacian: Demarcation of edges across multiple hierarchies, with greater prominence for the contour of the bird rather than higher hierarchical contours like those of the wing

(3) dx:



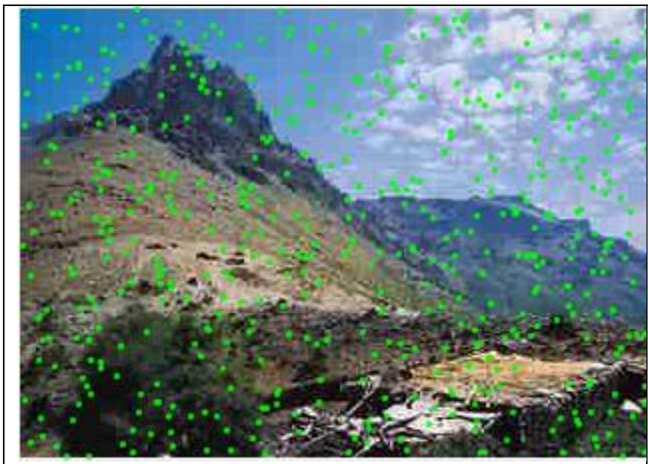
dX: Edge imprint with high preponderance of vertical or quasi vertical edges. The deeper the edges, the more vertical they are.

The CIE Lab color space is different from what I am use to on OpenCV as the channels are not distinct R, G, B channels but rather 2 of those channels are a blue à yellow and green à red, and the third one is a lightness channel (how bright the light is in the image or how dim etc). It is useful because it is device independent, which means that image processing should yield the same output for all devices irrespective of how color is defined on one machine. It is almost a version control for how images are defined by color, ensuring there is synergy and alignment on all devices that process those images.

Q1.2 Collect sample of points from image

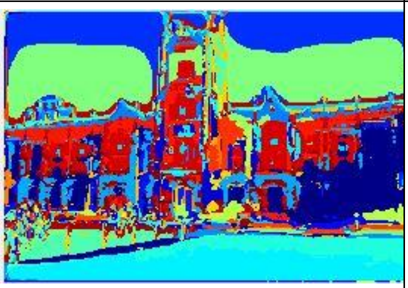
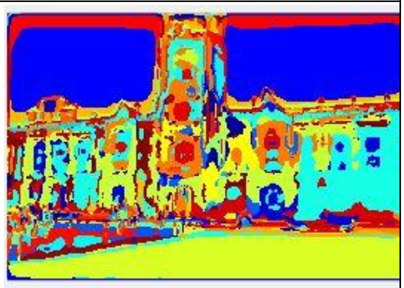
500-points collection

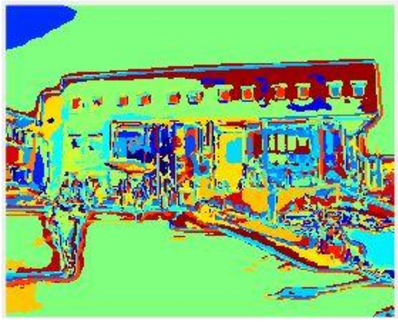
Random Points	Harris Points
	
	

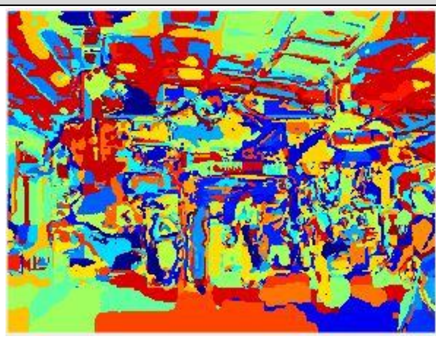
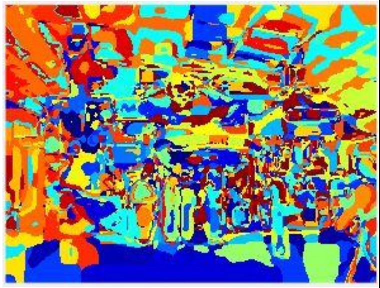
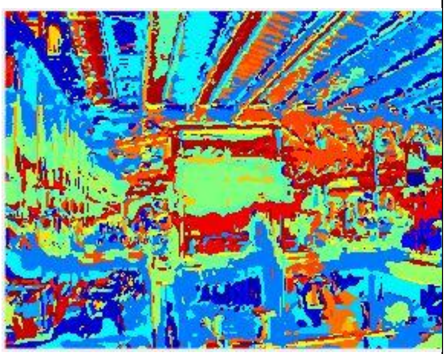
Q2.1 Convert image to word map

Class: Campus

Original	Harris	Random
		
		

Class: Airport

Original	Harris	Random
		
		
		

In general, the wordmaps are doing a decent job of capturing semantic meaning. There is a broad sense of categorization, however when looking a bit more closely, separate objects/regions are being represented by the same color. This looks to be especially true for the Harris word maps. A good example of this is the first image in the “Campus” table on the following page. The sky, part of the building, and the walkway all look to be the same shade of green despite being very separate things. Over this subset I would say that the random sampling produced better wordmaps. I believe this is because using Harris corner detection results in tight clusters of points. This is likely leading to the missing of important features that aren’t represented as corners. The random method samples uniformly so it avoids this clustering problem.

Q3.2 Evaluate Recognition System - NN and kNN

Harris - Euclidean	$\begin{bmatrix} 15 & 5 & 5 & 4 & 3 & 2 & 5 & 2 \\ 2 & 10 & 3 & 2 & 1 & 5 & 3 & 3 \\ 1 & 0 & 10 & 0 & 2 & 2 & 2 & 0 \\ 1 & 1 & 1 & 6 & 2 & 2 & 5 & 2 \\ 0 & 1 & 0 & 1 & 3 & 0 & 0 & 0 \\ 0 & 1 & 1 & 3 & 4 & 5 & 2 & 0 \\ 0 & 1 & 0 & 3 & 2 & 1 & 3 & 0 \\ 1 & 1 & 0 & 1 & 3 & 3 & 0 & 13 \end{bmatrix}$ Accuracy:41.53%
Random - Euclidean	$\begin{bmatrix} 10 & 5 & 4 & 3 & 2 & 4 & 4 & 1 \\ 4 & 10 & 1 & 1 & 4 & 3 & 3 & 3 \\ 2 & 2 & 9 & 3 & 2 & 3 & 3 & 0 \\ 1 & 0 & 3 & 8 & 2 & 3 & 2 & 1 \\ 0 & 0 & 1 & 0 & 4 & 0 & 0 & 0 \\ 0 & 2 & 1 & 3 & 0 & 6 & 2 & 1 \\ 1 & 1 & 0 & 1 & 3 & 1 & 4 & 0 \\ 2 & 0 & 1 & 1 & 3 & 0 & 2 & 14 \end{bmatrix}$ Accuracy:41.53%
Harris - Chi^2	$\begin{bmatrix} 15 & 4 & 6 & 1 & 0 & 2 & 2 & 2 \\ 2 & 11 & 2 & 1 & 3 & 3 & 1 & 1 \\ 1 & 2 & 11 & 2 & 3 & 2 & 5 & 0 \\ 0 & 0 & 1 & 9 & 0 & 2 & 3 & 1 \\ 0 & 2 & 0 & 1 & 12 & 1 & 3 & 0 \\ 0 & 1 & 0 & 1 & 0 & 7 & 1 & 1 \\ 0 & 0 & 0 & 5 & 2 & 2 & 4 & 0 \\ 2 & 0 & 0 & 0 & 0 & 1 & 1 & 15 \end{bmatrix}$ Accuracy:52.50%

Random - Chi²	$\begin{bmatrix} 12 & 4 & 2 & 2 & 1 & 3 & 2 & 3 \\ 3 & 12 & 3 & 1 & 3 & 3 & 1 & 1 \\ 1 & 2 & 13 & 3 & 3 & 5 & 1 & 0 \\ 1 & 0 & 1 & 9 & 1 & 0 & 3 & 2 \\ 0 & 1 & 1 & 0 & 8 & 0 & 3 & 1 \\ 0 & 1 & 0 & 1 & 0 & 7 & 2 & 0 \\ 1 & 0 & 0 & 3 & 4 & 1 & 7 & 0 \\ 2 & 0 & 0 & 0 & 0 & 1 & 1 & 13 \end{bmatrix}$ Accuracy:50.62%

After a fair bit of tuning, I was able to get both dictionaries to perform comparably. Given that the wordmaps based on the random dictionary seemed to do a better job of capturing sematic meaning I would have thought that it would perform better than the Harris dictionary. As can be seen from the table, using the Chi^2 distance metric resulted in ~10% better accuracy than Euclidean distance. This is because the Chi^2 metric weights the elements of the feature vector.

Specifically, features that have a high and similar frequency in both vectors will be calculated as “closer” than another feature that has a low but similar frequency across vectors.