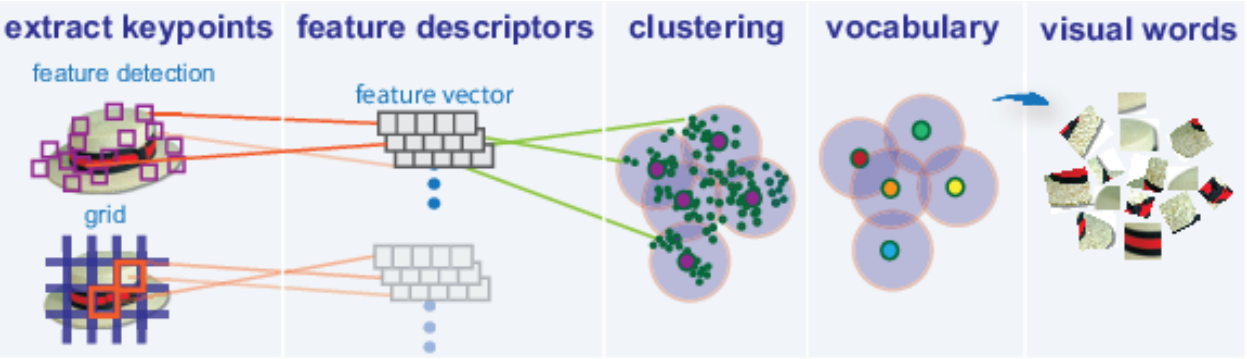
Assignment 3 – Bag of Visual Words

Name – Varsha Sureshkumar Maurya
NinerId – 801100600

In this assignment we will be implementing bag of visuals words and predict images based on the model.

Implementation Diagram: [1]

Bag of Words Implementation Diagram



Implementation detail

1. We first build a Bag of Visual Model using the following implementation.
    a. Extract SURF features for each of the training image. The result descriptor for each image are the key feature points in our algorithm. [M x 64]
    b. We run through all the training images and generate a big list of all the descriptors for all the images. [M * I * C * 64] where M is the total descriptor for each image. I the total number of images in each class and C is the total number of class that we have.
    c. Once, we have the total descriptor list with us, we use k-means clustering to cluster all the descriptors. K is a hyper-parameter which is chosen by tuning/ trail-and-error. The result will be K cluster with C centroid that will essentially be the visual words that we get.
    d. Lastly, the most important step is to generate a histogram of frequency of the descriptors of the image to find to which cluster are they closest to and build a frequency vector. We usually normalize this frequency vector so as to account in the factor of varying descriptors for each image. The result would be a vector with the frequencies of the descriptors w.r.t centers / Visual words. For example: If we have 8 descriptors for an image, we have total [8 x 64] descriptors and 5 Visual words meaning 5 centers, such that SURF descriptors of image are such that 5 are close to C=1, 2 to C=4 and 1 to C=5 the resultant V would be [5, 0, 0,2,1] respectively.
2. Thus, at the end of step 1 we will have created a Bag of visual word representation for our image set. We now use KNN to fit our model.
3. We validate it by using test images where we repeat the process 1.a and 1.d and use these as our features and predict the actual label.

Conclusion:

1. The total number of images used for training and building the bag of words model were N = 55 and C = 30, N=Number of images in each class, C= Total number of classes.
2. Hyperparameter
    a. K value for K-means:
    b. Total number of iterations for K-means:
    c. Number of Neighbors in KNN:
3. Prediction Accuracy:

Reference:
[1]: https://www.mathworks.com/help/vision/ug/image-classification-with-bag-of-visual-words.html
[2]: https://webpages.uncc.edu/kchinnak/SemiFinalProjectReport.pdf