GNR 638 Assignment 1 – Bag of Visual Words

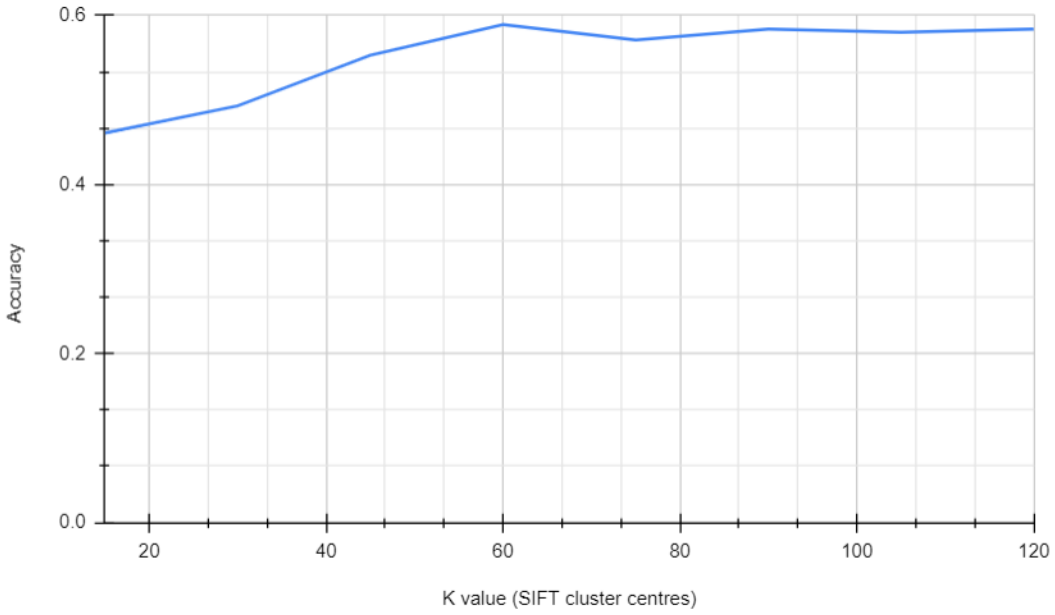
Amruta Mahendra Parulekar (20d070009)
Hemant Hajare (20d070037)

Question 1

Change K in the K-means part and see the accuracies. Make a graph of K vs Accuracy.

Experiment:

K was varied in steps of 15, up to 150 and the accuracy was plotted as follows:



Observation:

- The accuracy increases as the number of clusters increases and then it drops and stabilizes.

Conclusion:

- This can be explained by the fact that a very small number of clusters will put features of different objects into the same cluster (visual polysemy) and this will lead to more misclassification.
- As the number of clusters increases, features of distinct objects are put into distinct clusters and classification improves.
- If the number of clusters is increased further, very similar features may be unnecessarily split into different clusters (visual synonyms) and there is little improvement or even drop in accuracy. Thus, there is no benefit to the classification algorithm by increasing cluster number beyond a particular value, it will just unnecessarily increase computation.

Question 2

Add GLCM features (5 directions, four parameters per direction). Now, the feature length is K+20. Perform l_2-norm on the concatenated features before classification. See the performance.

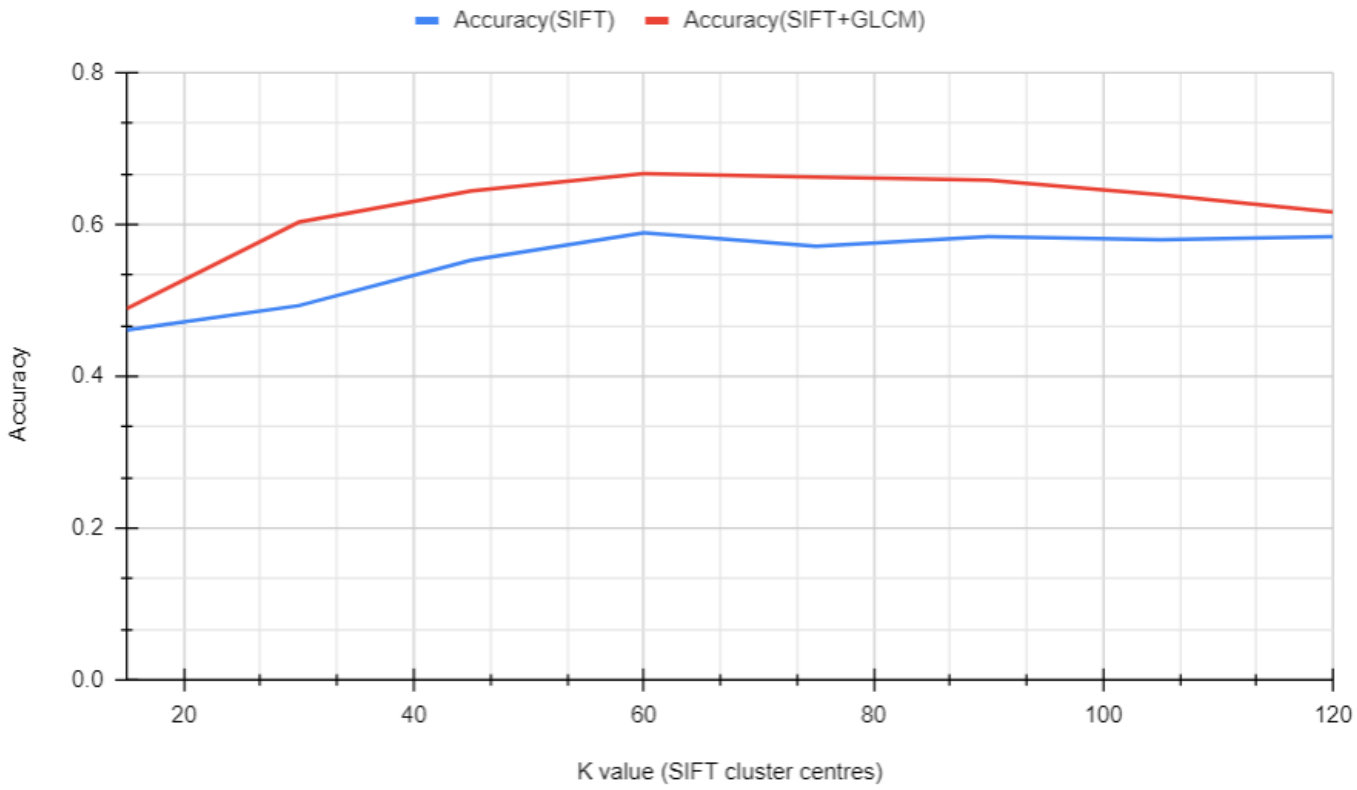
Experiment:

- We created a function that computes the GLCM matrix of every image. For this, we used the graycomatrix function of the skimage module.
- We used four directions for the GLCM and specified those using angles, namely 0, 0.7854, 1.5708 and 2.3562 radians. Distance for the GLCM was chosen based on resolution of the images, and we chose it to be 4 pixels
- Then we computed five major properties from each GLCM, namely contrast, dissimilarity, homogeneity, correlation and energy. These properties were calculated using the graycoprops function of the skimage module.
- 'contrast': $\sum_{i,j=0}^{levels-1} P_{i,j}(i-j)^2$
- 'dissimilarity': $\sum_{i,j=0}^{levels-1} P_{i,j}|i-j|$
- 'homogeneity': $\sum_{i,j=0}^{levels-1} \frac{P_{i,j}}{1+(i-j)^2}$
- 'ASM': $\sum_{i,j=0}^{levels-1} P_{i,j}^2$
- 'energy': $\sqrt{ASM}$
- **'correlation':**

$$\sum_{i,j=0}^{levels-1} P_{i,j} \left[\frac{(i-\mu_i)\,(j-\mu_j)}{\sqrt{(\sigma_i^2)(\sigma_j^2)}} \right]$$

- Thus, we now had 5 parameters in 4 directions, each could be considered a feature of the image.
- Thus, we had 20 glcm features which could now be used for classification.
- Due to the bag of visual words algorithm, each image had been represented using k numbers, each number telling you how often a SIFT feature from a particular cluster (out of k clusters) occurs in the image
- After standardization of the BoVW and GLCM features separately, we concatenated the 20 new numbers to the k original numbers and finally normalized all features together using l2_norm

- Then we used the existing SVM classifier to get an improvement in classification accuracies as shown in the following plot:



Observation:

- Addition of GLCM features to SIFT features leads to an improvement in the classification accuracy
- The accuracy increases as the number of clusters increases and then it drops and stabilizes.

Conclusion:

- Addition of GLCM features leads to texture and pattern recognitiona and incorporation in the classification algorithm.
- Thus, it improves the ability of the SVM to classify by giving it additional information about the objects in the images
- A very small number of clusters will put features of different objects into the same cluster (visual polysemy) and this will lead to more misclassification.
- As the number of clusters increases, features of distinct objects are put into distinct clusters and classification improves.
- If the number of clusters is increased further, very similar features may be unnecessarily split into different clusters (visual synonyms) and there is little improvement or even drop in accuracy. Thus, there is no benefit to the classification algorithm by increasing cluster number beyond a particular value, it will just unnecessarily increase computation.

Question 3

Read about Visual Polysemy and Synonyms in the context of BoVW. Note they are also in the slide in Shared. Check manually if these occur in the data shared on the git page.

- **Visual Polysemy:** Single visual word occurring on different (but locally similar) parts on different object categories.
- For example, in images such as the one below from the Git dataset, the annotated SIFT features belong to different objects (cupboard and sofa) but they are locally similar and will be grouped into the same cluster by the algorithm

- **Visual Synonyms:** Two different visual words representing a similar part of an object (wheel of a motorbike)
- For example, in images such as the one below from the Git dataset, the annotated SIFT features belong to a similar part of the same object (centre of cloud) but they are locally different and will be grouped into different clusters by the algorithm