Constrained Clustering of Images with Textual (or Visual) Explanation

Advanced Topics in Machine Learning- Semester Project, 2022

Motivation

The PASCAL Visual Object Classes Challenge (VOC 2007) dataset contains images of various objects and classes. With the help of clustering, similar images can be grouped together, which makes it easier for searching and navigation.

The clustering depends on the type of feature extraction technique used. Once done, a user will be able to perform clustering on a smaller subset of images based on a query.

The dataset contains over 5000 images. Each image may contain objects from multiple classes.

Data, Constraints Creation & Constraints Clustering

Dataset: The PASCAL Visual Object Classes Challenge (VOC 2007) data set consists of about 5000
images of 20 classes. The twenty object classes that have been selected are:

- person
- bird, cat, cow, dog, horse, sheep
- aeroplane, bicycle, boat, bus, car, motorbike, train
- chair, dining table, potted plant, sofa
- tv monitor

Constraints Creation: A distance matrix is created using the candidate images (the images that have been selected using kNN for a given query image). The number of must-link constraints are set to the number of object classes found in the candidate images. To create the contraints, an image is selected at random from the candidate images, and a must-link is formed with it's nearest neighbors using 2NN.

Constraints Clustering: Used algorithms are COP-K-means and PCK-means. COP-K-Means demonstrates clustering with hard constraints and PCK-means demonstrates clusters with soft constraints.

Feature Extraction

1. Bag of Visual Words (BOVW)

The general idea of bag of visual words (BOVW) is to present an image as set of features. Features consists of keypoints and descriptors.

The steps involved are as follows-
Extracting features locally - Obtaining features of LBP for each patch
Cluster the local features - Using the clustering algorithm, KMeans, defining the number of visual words
Histograms from Bag of features - Checking the frequency of each visual word in each training image to form the histogram

2. MPEG7 Color Layout Descriptor

It is designed to capture the spatial distribution of color in an image. The feature extraction process consists of two parts: grid based representative color selection and discrete cosine transform with quantization.

The steps involved are as follows-
Image partition - Divide the image into 64 blocks, a grid of 8x8

Representative color selection - A single color is selected as the representative of that block from each block.

DCT transformation is applied to each matrix of Y, Cb and Cr components for each block.

3. Scale-invariant feature transform (SIFT)

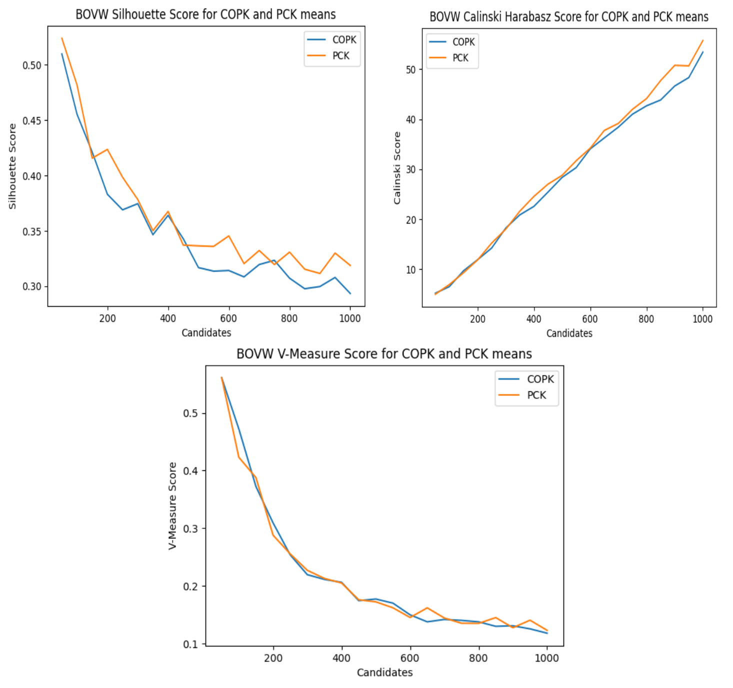
SIFT helps locate the local features in an image, commonly known as the 'keypoints' of the image.

The steps involved are as follows-
Formation of Scale space
Detection of Keypoint using Difference of Gaussians (DoG)
Estimation of keypoint orientation using image gradients and direction histograms
Extraction of description of keypoints

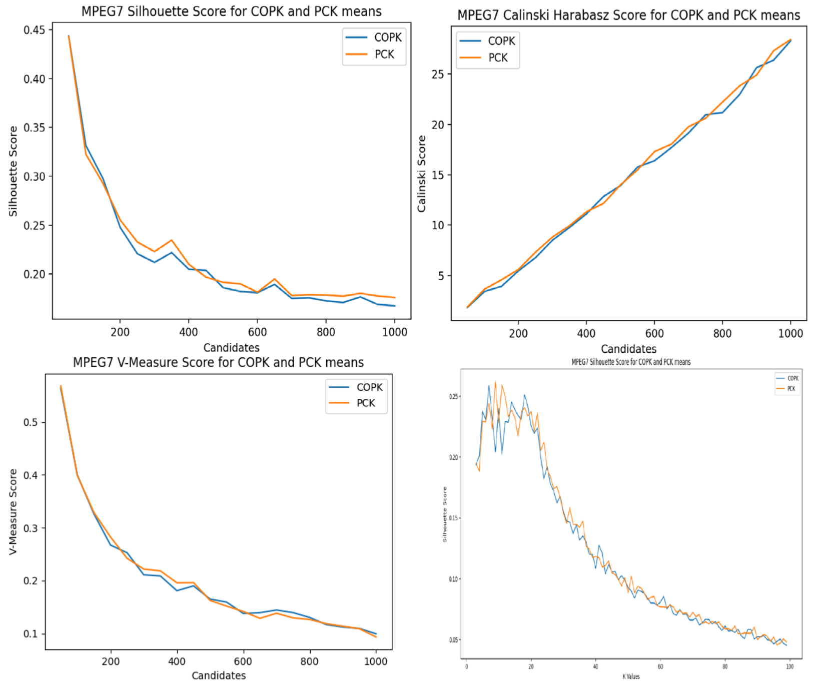
Evaluation

To evaluate performance of clusters on COPK and PCK means, the following measures are used-
Silhouette Score: In silhouette score, data set is used for measuring the mean of the Silhouette Coefficient for each sample belonging to different clusters.
V-Measure: It is defined as the harmonic mean of homogeneity and completeness of the clustering.
Calinski-Harabasz Index: It is defined as the ratio between the within-cluster dispersion and the between-cluster dispersion.
Query Image: "000036"

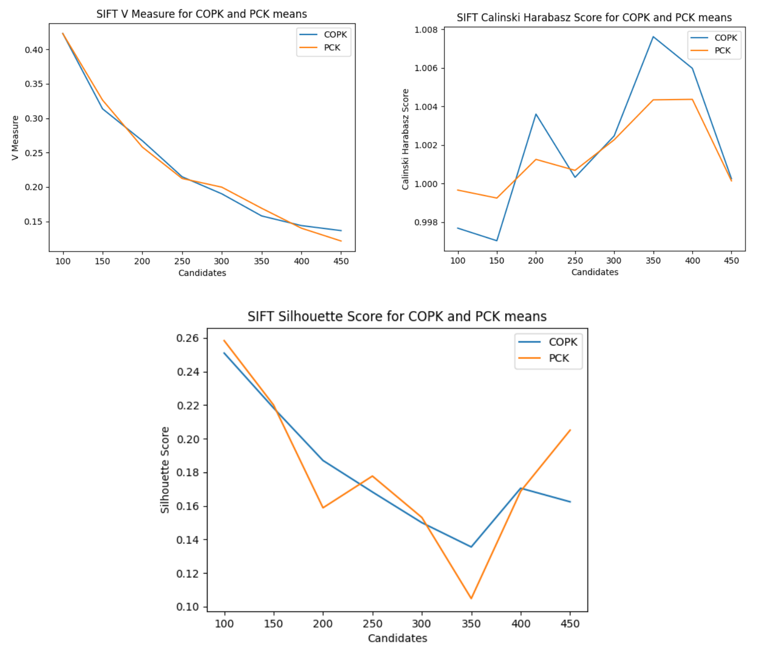
BOVW Evaluation



MPEG7 Evaluation



SIFT Feature Evaluation


SIFT V Measure for COPK and PCK means
V Measure
Candidates
COPK
PCK
SIFT Calinski Harabasz Score for COPK and PCK means
Calinski Harabasz Score
Candidates
COPK
PCK
SIFT Silhouette Score for COPK and PCK means
Silhouette Score
Candidates
COPK
PCK