Constrained Clustering of Images with Textual (or Visual) Explanation

Advanced Topics in Machine Learning- Semester Project, 2022

**Dataset:** PASCAL Visual Object Classes Challenge (VOC 2007) data set which consists of about 9000 images and 20 classes.

**Constraints Clustering:** Used algorithms are COP-K means and PCK means.

**Data and Constraints Clustering**

**Motivation**

**2. Moving Picture Experts Group (MPEG7)**

MPEG-7 is a standard for describing features of multimedia content. The goal of the MPEG-7 standard is to provide a rich set of standardized tools to describe multimedia content.

The steps involved are as follows-

**Read the image using OpenCV's** imread method, which creates a numpy array from the image.

**Image partition** - Divide the image into 64 blocks, a grid of 8x8. The 8x8 grid ensures invariance to resolution or scale.

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

Get a tiny image of 8x8 in YCbCr color space by converting color space from RGB to YCbCr.

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

Scan each matrix in a zigzag fashion to group the low-frequency coefficients of the matrices.

**3. Scale-invariant feature transform (SIFT)**

SIFT helps locate the local features in an image, commonly known as the 'keypoints' of the image. These keypoints are scale & rotation invariant that can be used for various computer vision applications, like image matching, object detection, scene detection, etc.

The steps involved are as follows-

**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. Keypoints are “stand out” points in an image and descriptor is the description of the keypoints.

The steps involved are as follows-

**Extracting features locally:** defining the size of the patch and the number of patches per image and using LBP features as a base descriptor. 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:** check the frequency of each visual word in each training image. First, computing features for each image and predicting the number of patches of images. Finally, computing histogram and appending in the array.

**Feature Extraction**

**Evaluation**

To evaluate clusters,