Date – 5th February 2019

CS783

Assignment 1

Instance Recognition

Team Members

Mohammad Areeb (160403)

Somesh Pandey (160699)

Training Dataset Pre-processing

1.The chessboard present in the training dataset was posing a problem as SIFT and SURF were extracting too many features from it and the final ranking of images was seriously affected by this.

2.For example, a test image containing 'listerine' and having a chessboard in a certain orientation was having maximum matches with all training images having the chessboard in that orientation, giving very less importance to the features extracted from 'listerine' itself.

3.Hence, in order to remove the impact of chess board, we cropped the training dataset images using OpenCV, and then used this cropped dataset for further training.

Methods used for Feature Extraction

1.Scale Invariant Feature Transform (SIFT)

2.Speeded Up Robust Features (SURF)

3.Combination of the above two

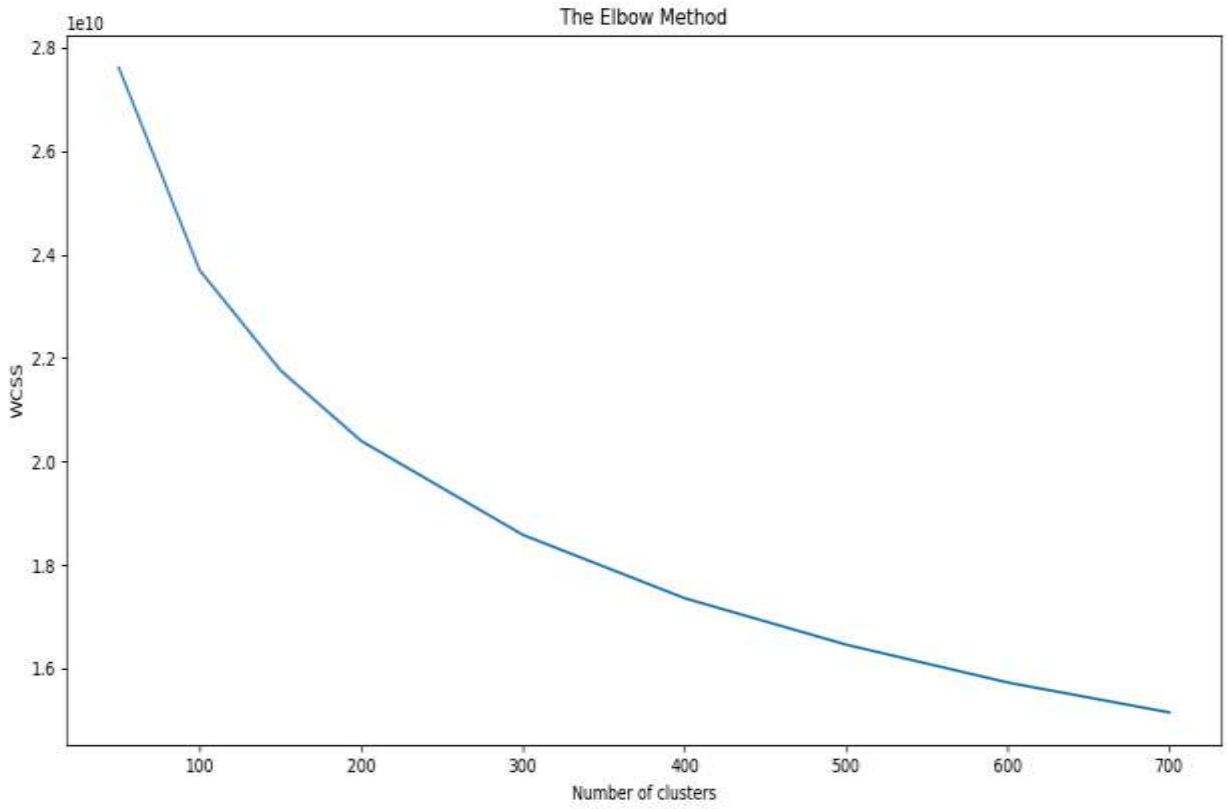
We have used OpenCV library in python for implementing SIFT and SURF features from training dataset.

We tried working with only SIFT features, only SURF features and also tried a combination of SIFT and SURF.

Clustering Algorithm

We have used K-Means clustering algorithm for finding the visual words.

We tried for k = 50, 100, 150, 200, 300, 400, 500, 600, 700. Elbow plot is attached →



We further tried for larger values of K. However, the process was time consuming and computationally expensive.

Hence, we switched to a faster approach → Mini Batch K-Means Clustering and moved the training part to 'Google Colab' for using online-GPU resources.

We finally used k = 10000 for clustering.

Scoring and Ranking

We calculated the TF-IDF scores of training data and test data and used the cosine similarity criteria for calculating the similarity score.

Based on the scores, we sorted the images for each query and stored it.

Problems Faced

1.As mentioned above, first problem was the overemphasis to the features extracted from the chessboard which led to severe discrepancies in results.

2.Very slow convergence of the K-Means algorithm and lack of processing power.

3.Finding optimum value of 'k' for k-means clustering.

4.Since several advanced Instance Recognition approaches were available online, and the time and resources were limited, we couldn't give it our best.

Scope for Improvement

We plan to use YOLO (You Only Look Once) object detection algorithm to find the bounding boxes for all images in the training dataset and crop them accordingly to reduce the impact of chessboard and other unwanted features like table edge, etc.

We hope to achieve high accuracy using these newly extracted features.

We also plan to test Faster-RCNN methods which mainly employ Content Based Image Retrieval (CBIR) techniques.