CSE 515 (Fall2016) Phase 3 Report: Group 17

Anshuman bora, Arizona State University

vraj delhivala, Arizona State University

satyam jaiswal, Arizona State University

saumya parikh, Arizona State University

Tanmay patil, Arizona State University

**Abstract**

Finding similarity of frames between multiple videos is not enough. They also have to be listed according to some kind of importance or relevance. In phase2, we listed them by sorting with respect to the similarity index. The frame with highest similarity came first and the top *k* frames were selected as most similar sub-sequence. In phase3, we have improved the listing logic by implementing intelligent algorithms which list similar frames by significance, relevance and also by nearest neighbor importance in high dimensional spaces. SIFT vectors in a reduced space have been used as input considering they describe important visual features. Algorithms like PageRank and ASCOS++ have been used for ranking according to significance. Relevance feedback based algorithms like personalized PageRank and personalized ASCOS ++ have been used and nearest neighbor based technique of Locality Sensitive Hashing has also been used. This report talks about all these implementations in detail.

**Keywords**

**Feature Extraction, Similarity Graph, Significant Frame, Relevant Frame, Page Rank, ASCOSS, Personalized PageRank, Locality Sensitive Hashing, Multi-dimensional Indexing**

# Introduction

Terminology

TODO.

Goal Description

Phase 3 involves around experimenting with clustering, indexing, classification and relevance feedback. The detailed goal description are as follows:

1. Task1: In task1 we had to implement Video Feature Extraction algorithm where Principal Component Analysis (PCA) had to be applied to the original space to get a reduced vector space.
   1. Input: The input to this algorithm is ‘*d*’ (the target reduced dimensionality) and *‘filename.sift’* (file containing SIFT features)
   2. Output: the output would be a file named ‘filename\_d.spc’ which would contain entries in the form of *{<i, j, l, x, y>, [dim-1 , . . , dim-d]}*, such that i = video, j = frame and l = cell indices. x and y and provide the position of the SIFT keypoint and [dim-1, . . , dim-d] are the reduced dimensions. The program is also expected to output the d dimensions in the form of *<original\_index, score>* in non-increasing order of scores.
2. Task2: In task2 we have to create a similarity graph using the reduced SIFT vector space.
   1. Input: Integer *K* and SIFT keypoints in file *‘filename\_d.spc’.*
   2. Output: Graph(V, E) where V are the nodes corresponding to each frame and E are the edge pairs corresponding to <va, vb> where vb is one of the most similar frames to vb, from a different video that vb. This graph will be stored in *‘filename\_d\_k.gspc’* in the format of *<va, vb, sim(a,b)>*
3. Task3: The goal of task3 is to list most similar frames according to significance. This is to be done using 2 methods namely, PageRank and ASCOS++.
4. Task4: This task is similar to the previous task. The only difference is that the similar frames are to be listed according to relevance. The algorithms to be used are Personalized PageRank and Personalized ASCOS++.
   1. Input: Filename *‘filename\_d\_k.gspc’* and integer *‘m’* where m is number of the most significant frames to be listed.
   2. Output: Visualize the *m* frames for PageRank and ASCOS++ both.
5. Task5: This task involves implementing multi-dimensional index structures and nearest neighbor search using Locality Sensitive Hashing (LSH) tool.
   1. Input: Filename *‘filename\_d.spc’* , number of layers *L*,the number of buckets *2^k.*
   2. Output: File *‘filename\_d.lsh’* containing entries in the form {layer\_num, bucket\_num, <i;j;x;y>}.
6. Task6: A similarity-based video object search tool is to be implemented.
   1. Input: LSH index file, integer n, object in the form *{i ,j, <x1, y1>, <x2, y2>}* such that <x1, y1> & <x2, y2> is a rectangle containing the object.
   2. Output: Visualize the top n frames. Also output should contain:
      1. Number of unique SIFT vectors considered.
      2. Number of overall SIFT vectors considered.
      3. Number of bytes of data from the index accessed to process the query.

Assumptions

For this phase, we have made the following general and specific assumptions:

* Overall: Nice
* Task1: Nice
* Task3: Nice
* Task4: Nice
* Task5: Nice
* Task6: Nice

# implementation

Task 1: Video Feature Extraction

Nice.

Task 2: Video Frame Similarity Graph Generation

Nice.

Task 3: Most Significant Frame Selection

Nice.

Task 4: Most Relevant Frame Selection

Nice.

Task 5: Multi-dimensional Index Structures and Nearest Neighbor Search

Nice.

Task 6: Similar Video Object Search

Nice.