Storyline Reconstruction for images

Interim Project Report

Arpit Khandelwal

Sameedha Bairagi

Venkatesh Raizaday

Introduction:-

Storyline reconstruction is a relatively new topic in which the main task is to take a stream of images as an input and sort them in a chronological order. The usual strategy is to get an understanding of temporal dependencies of an activity using a source like videos. Using this learning, we try and classify the images in the order of time. So basically, it is an intelligent sort. ☺

We have restricted the categories of videos to the following: High Jump, Shot put, Disc Throw, Baseball pitch, Basketball, Field Hockey Penalty, Blowing Candles, Bowling, Clean and Jerk, Cutting, Diving, Golf Swing, Javelin Throw, Long Jump, Pole Vault, Skiing, Soccer Penalty, Tennis swing, Volleyball Spiking.

These categories were chosen for these reasons:-

- Low repetition of activities.
- Well defined distinct key frames.

The videos have been taken from the **UCF101** dataset which is available for free on http://crcv.ucf.edu/data/UCF101.php .

Background and Related Work:-

The topic is relatively new and not much work has been done on it. Most of our understanding of the topic has been taken from **Joint Summarization of Large-scale Collections of Web Images and Videos for Storyline Reconstruction by Kim, Sigal and Xing.**

In this paper, for each video the author finds the k – nearest photo streams using naïve bayes nearest neighbor. Then, builds a similarity graph between the video and stream of images. Then, the author optimizes the graph by focusing on densely connected nodes and choosing nodes that are distant from one another. The feature space used is color SIFT and HOG features.

Progress so Far:-

We have decided to implement a homegrown unsupervised algorithm for the problem. For each video in the database, we generate multiple frames from the video. Then we try to create a video summary of key frames. Currently, we have two hypothesis on how to generate key frames:-

- Find the key frames of each video by using k – means algorithm on the HOG features of the all the frames of that video. After generating k clusters, merge the kth cluster of all the videos to give kind of universal clusters.
- Get all the frames from all videos and generate k clusters from all the images.

Both these approaches have theoretical pros and cons and we will experiment using both to find the best results.

Now these clusters are arbitrary and to generate the temporal dependency of these clusters, we compare these to a video from the training set. The clusters will ideally have minimum distance from one of the frames from the video. The ordering of the clusters are deduced from the ordering of the frames from the video.

Now when a new set images is given for classification, we convert the images to HOG features and calculate the distance of each image from the key frames. Every image is assigned the cluster with minimum distance and classified based on the temporal state of the cluster.

The dataset contains 19 categories with each category having 100 videos. We have generated the frames for each video and are currently experiment on trying to find a working configuration for generating desired clusters.

Trying to use HOG features or SIFT has given us arbitrary clusters, hence we are also focusing on different features specially features based on optical flow.

Also if time permits, we also intend to convert the problem statement into a classification problem.

Revised Research:-

- 04/10: Compile a list of feature vectors to experiment with.
- 04/15: Complete clustering and feature selection.
- 04/17: Complete classification and record accuracy.
- 04/20: Formalize an approach for classification based storyline construction.
- 04/22: Complete the poster.
- 04/27: Finish coding for supervised approach.
- 04/30: Summarize results and finish report.

References:-

- Joint Summarization of Large-scale Collections of Web Images and Videos for Storyline Reconstruction
- G. Kim and E. P. Xing. Jointly Aligning and Segmenting Multiple Web Photo Streams for the Inference of Collective Photo Storylines. In CVPR, 2013
- A. Khosla, R. Hamid, C. J. Lin, and N. Sundaresan. Large-Scale Video Summarization Using Web-Image Priors. In CVPR, 2013.
- H. Misra, F. Hopfgartner, A. Goyal, P. Punitha, and J. M. Jose. TV News Story Segmentation Based on Semantic Coherence and Content Similarity.