Video Summarization

Animesh Ramesh (14511), **Archit Sharma** (14129), **Kanishk Gandhi** (14235), **Nikhil Vanjani** (14429), **Shibhansh Dohare** (14644)

Motivation

Rapid growth in digital media has led to an outburst of videos. Digital content, especially videos, have become easier to generate than consume. This warrants the need for automatic video summarization. The video summarization models can be used to develop trailers for movies and TV Shows, generate highlights for sports matches or extract important events from surveillance data.

Task Description

The basic unit of videos is a frame, which is a 2D collection of RGB colors. These frames when viewed at a rate beyond cutoff (Typical frame rate for videos is 30 frames per second), would generate an effect of continuity. Video summarization aims at creating a summary with maximum amount of information with least amount of frames, which essentially means filtering the critical frames (or sequences of frames) from the original video.

Models and Techniques proposed

Video Summarizaration is typically treated as an unsupervised learning problem. However, some recent work models video summarization as a supervised learning problem. Typical hierarchy for the videos is scenes, then shots, and then frames. Summarization can be accomplished using two techniques: **Keyframe or static video summarization**, which outputs the "key" frames from the videos and **Video skimming or dynamic video summarization**, which outputs a collection of shorter clips from the video.

For static video summarization, following methods will be tested subject to time constraints

- Shot boundary based
- Clustering based methods for choosing frames (k-means, hierarchical)
- Perceptual feature based (color histogram, motion-based and object based)
- Semantic Feature based keyframe selection
- Scene-Change Detection

For dynamic video summarization, following methods will be tested

- Model-Based Summarization
- Time-Compression based
- Text and speech recognition based

Effect of dimensionality reduction techinques such as PCA and SVD, and various transforms such as Discrete Haar Wavelet transforms, SIFT (Scale Invariant Feature Transform) are also proposed to be studied, wherever relevant. Supervised video summarization may also be explored.

Dataset

Many datasets are available publicly (Example LEAR). Other datasets will be explored as well.

References

1. Video Summarization: Techniques and Applications - *Zaynab El khattabi, Youness Tabii, Abdelhamid Benkaddour*
2. Diverse Sequential Subset Selection for Supervised Video Summarization - *Boqing Gong, Wei-Lun Chao, Kristen Grauman, Fei Sha*
3. Video Summarization - *Ben Wing*