

A Multi-Faceted Video Summarization System: Tradeoff Between Diversity, Representation, Coverage and Importance

Vishal Kaushal, Rishabh Iyer, Khoshrav Doctor, Anurag Sahoo,
Pratik Dubal, Suraj Kothawade, Rohan Mahadev,
Kunal Dargan, Ganesh Ramkrishnan

Motivation

People remember...



20%

20% of what they hear



30%

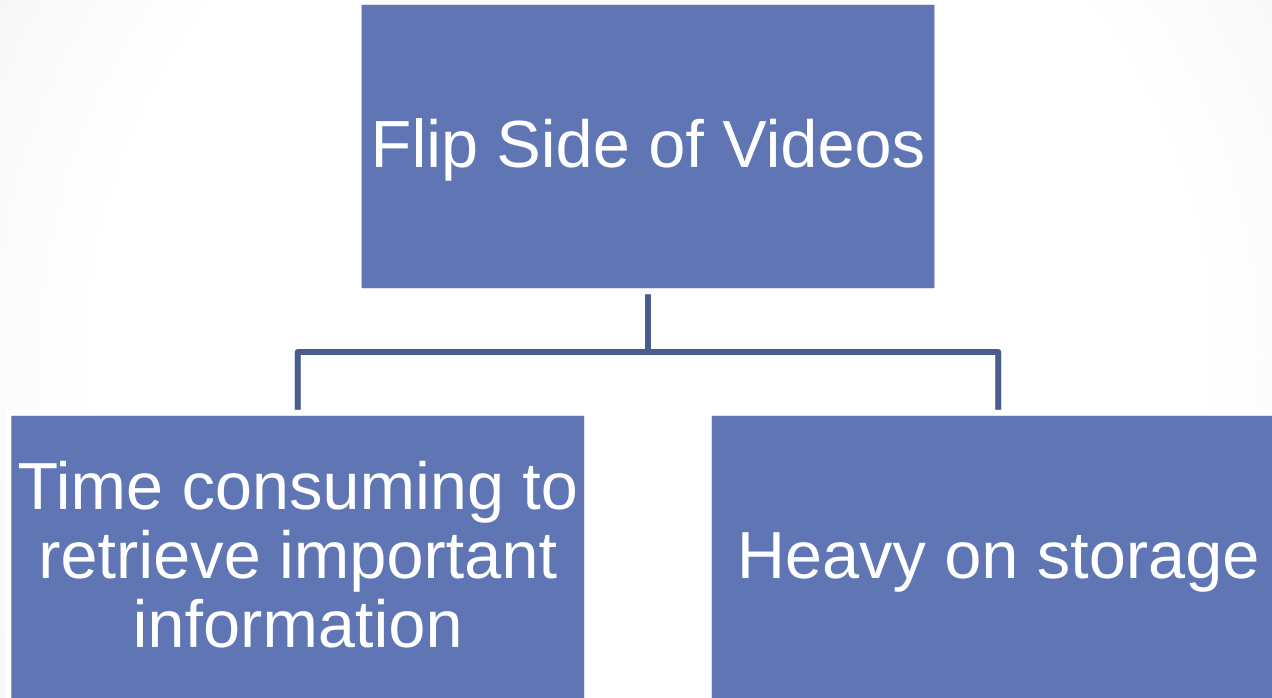
30% of what they see



70%

An amazing 70% of
what they see & hear!

Motivation



Motivation

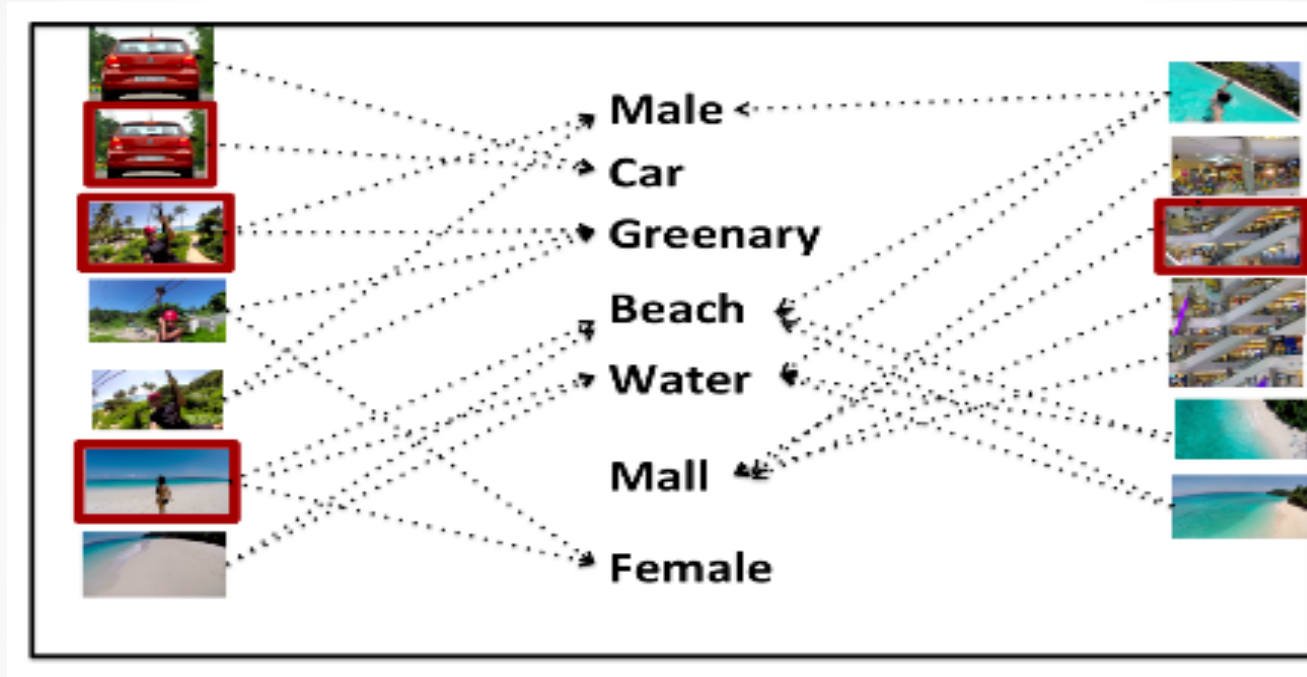
- Growing focus on Video Summarization
- Multi-faceted video summarization – Extractive, Query-based, Entity or Concept based
- Explainability of different models and featurizations and how they apply to different applications
 - Submodular functions naturally model notions of diversity, coverage, representation, importance
 - For e.g., in a surveillance video, diversity is more important, while in a movie, coverage and representation becomes more important
- Optimizations and tricks essential for building a large scale real world video summarization system

Diversity Functions



E.g. Dispersion (Disparity) functions, Determinantal Point Processes

Coverage Functions



E.g. Set Cover Function, Probabilistic Set Cover,
Feature Based Functions

End-to-end video summarization system

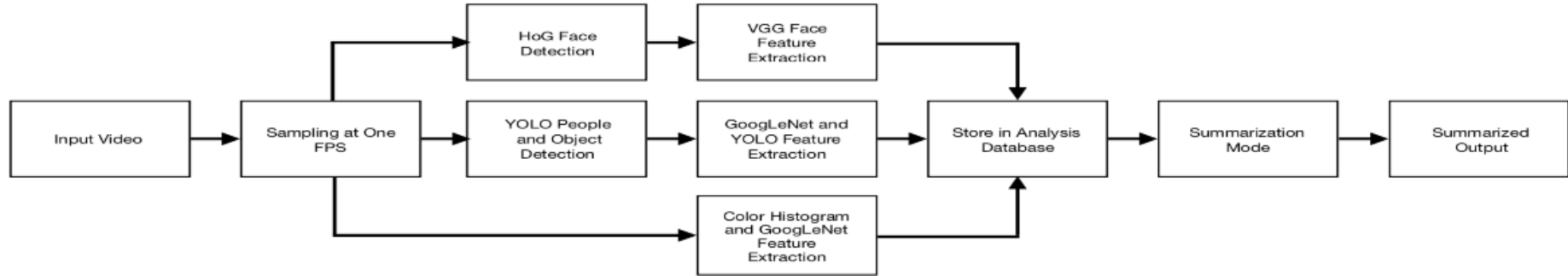


Figure 4. End-to-End Processing and Summarization of a Video

Results

Extractive Summary: TV Shows (Friends, Season 1, Episode 1)

Diversity Function (Disparity Min)



Coverage Function (Set Cover)



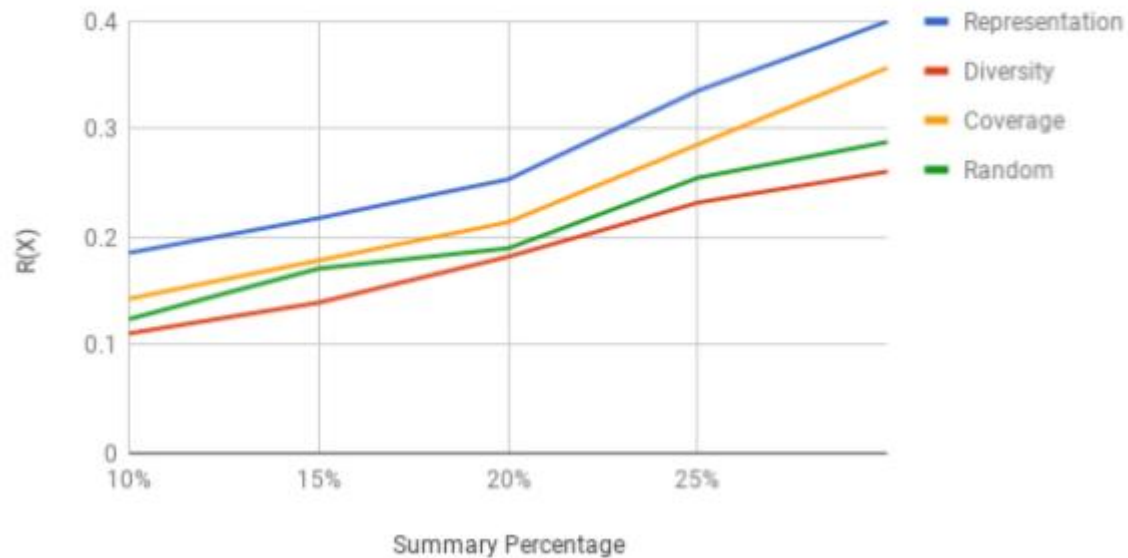
Representation Function (Facility Location)



Results

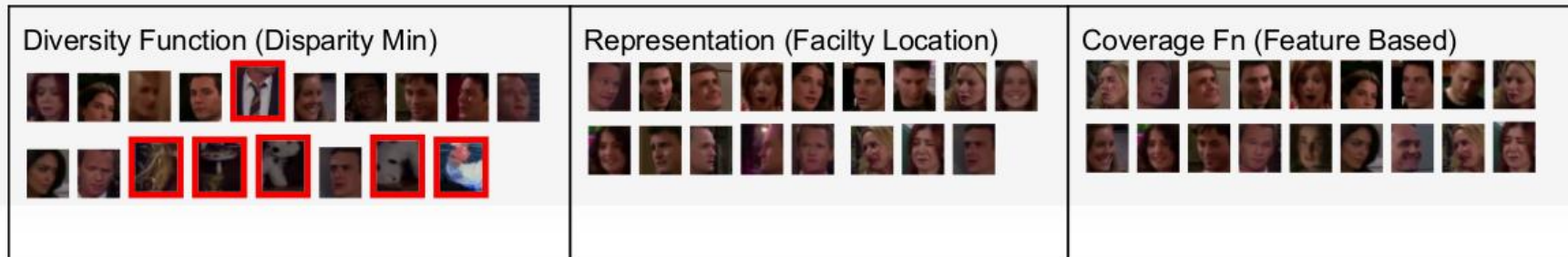
Extractive Summarization: Representation

TV Show and Movie Videos



Results

Entity Summary (Faces): TV Shows (How I Met Your Mother)



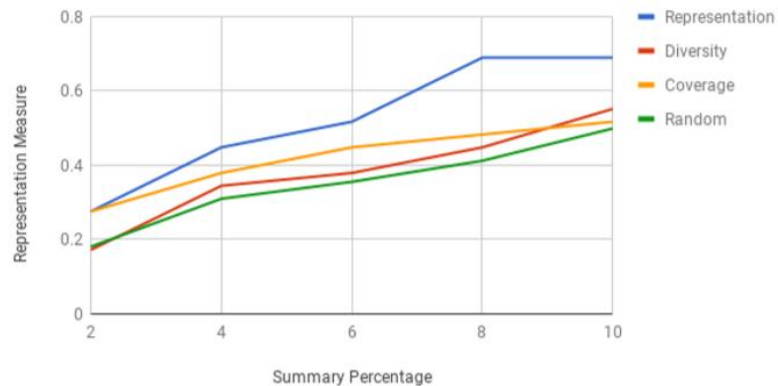
Query Based Summary (SkyScrapers): Travel Video



Results

Concept Summarization (Objects): Representation

Set of Object Entities



Query Summarization (Scenes): Diversity

Travel Videos

