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% of what they hear

30% of what they see

70% of what they see & hear!
Motivation

Flip Side of Videos

- Time consuming to retrieve important information
- Heavy on storage
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
  o Submodular functions naturally model notions of diversity, coverage, representation, importance
  o 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
Representation Functions

E.g. Facility Location Function, Saturated Coverage Function, Graph Cut Functions
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
Extractive Summary: TV Shows (Friends, Season 1, Episode 1)

<table>
<thead>
<tr>
<th>Diversity Function (Disparity Min)</th>
<th>Coverage Function (Set Cover)</th>
<th>Representation Function (Facility Location)</th>
</tr>
</thead>
<tbody>
<tr>
<td>![Images of Friends scenes]</td>
<td>![Images of Friends scenes]</td>
<td>![Images of Friends scenes]</td>
</tr>
</tbody>
</table>
Results

Extractive Summarization: Representation
TV Show and Movie Videos

- Representation
- Diversity
- Coverage
- Random

Summary Percentage

R(X)
## Results

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

<table>
<thead>
<tr>
<th>Diversity Function (Disparity Min)</th>
<th>Representation (Facility Location)</th>
<th>Coverage Fn (Feature Based)</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Images" /></td>
<td><img src="image2.png" alt="Images" /></td>
<td><img src="image3.png" alt="Images" /></td>
</tr>
</tbody>
</table>

### Query Based Summary (Skyscrapers): Travel Video

<table>
<thead>
<tr>
<th>Diversity Function (Disparity Min)</th>
<th>Representation (Facility Location)</th>
<th>Coverage Fn (Feature Based)</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image4.png" alt="Images" /></td>
<td><img src="image5.png" alt="Images" /></td>
<td><img src="image6.png" alt="Images" /></td>
</tr>
</tbody>
</table>
Results

Concept Summarization (Objects): Representation

Query Summarization (Scenes): Diversity