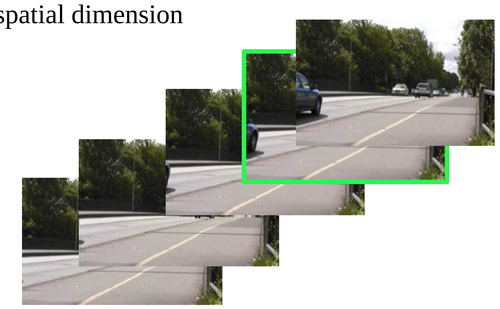
Temporal Saliency in Videos

Nicholas Baker, Nasim Souly
(nbaker9@jhu.edu, nasim.souly@gmail.com)
University of Central Florida

1. Problem

- **Objective:**
  - Identify frames of semantic interest in video clips with methods used in locating spatially salient regions
- **Applications:**
  - Video summarization
  - Improved understanding of what makes something salient along the spatial dimension



2. Data Set

- Used combination of Youtube videos and academic test videos
- Static camera images
- Surveillance and traffic monitoring videos

Sample Surveillance Video from Youtube

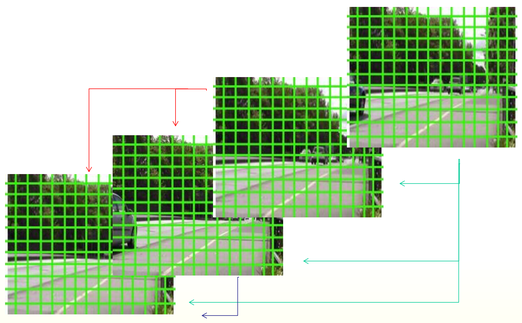


3. System Pipeline

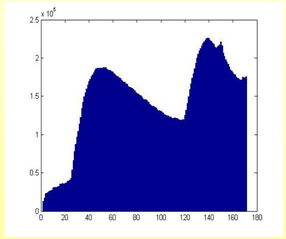
Input Video

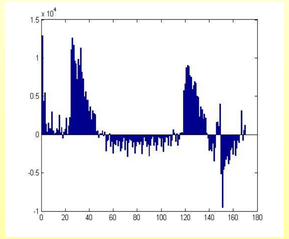


For each frame *i*, analyze dissimilarity *i* to all posterior frames



Compute derivative of dissimilarities and extract significant frames









⋮

Dissimilarity Calculation

$$\text{Dissim(current frame)} = \frac{\sum_{k=1}^{f}\left(framedif(\boldsymbol{f},k)*\frac{k}{f}\right)}{\boldsymbol{f}}$$

Where f is the total number of frames preceding the current frame

$$\text{Framedif(a, b)} = \sum_{i=1}^{w} norm(a.window(i) - (b.window(i))$$

Where w is the total number of windows in either a or b

Grouping

- One event will often trigger several points of high activation
  - Select only first frame and ignore subsequent frames until activation returns to < 0
  - Or take all frames to get a clip of temporally salient event

Selecting frames

- Thresholding
  - Normalize derivative of dissimilarity and select frames above chosen value (e.g., > .8)
- Winner take all
  - Find single frame with greatest temporal saliency

4. Results

- Tested on four video clips
- In all clips, there was a large increase in temporal saliency in frame where a semantically interesting event occurred
- System might be overly sensitive to light changes that are not semantically interesting

