Story-Driven Summarization for Egocentric Video

Zheng Lu and Kristen Grauman
University of Texas at Austin

Our idea
Produce a short visual summary that contains the story of an egocentric video.

Subshot selection objective
Given a series of subshots segmented from the input video, our goal is to select the optimal K-node chain of subshots (or keyframes):

\[ S^* = \arg \max_S Q(S) \]

\[ Q(S) = \lambda_x S(S) + \lambda_y I(S) + \lambda_d D(S) \]

Story progress between subshots
Good story: a coherent chain of subshots, where each strongly influences the next.

Story-driven spectral clustering

Potential applications
- Memory aid
- Law enforcement
- Mobile robot discovery

Temporal subshot segmentation
We propose a temporal segmentation method tailored to egocentric video that detects core ego-activities.

Subshot and object representation
For egocentric video, the story of activity is largely defined by objects.

Coherent object activation patterns

Results

Datasets
- UT Egocentric (UTE) (Lee et al. CVPR 2012)
- 4 videos, each 3-5 hours long, uncontrolled environment.
- We use visual words and subshots.

Evaluating summary quality
Large-scale user study:
- UTE: 6 videos and 33 events.
- 34 subjects, from 18 – 60 years old.
- ADL: 7 hours and 37 events for ADL.
- 5 users per comparison. Total 535 tasks, 45 hours of user time.

Evaluating summary quality

Activities of Daily Living (ADL) (Pichavant & Ramanan CVPR 2012)
- 20 videos, each 20-60 minutes, daily activities in house.
- We use object bounding boxes with keyframes.

Evaluating summary quality

UTE example
Our method

ADL example
Our method

Discovering influential objects
- Ours: objects ranked by influence.
- Baseline: objects ranked by frequency.
- Ground truth: workers on MTurk identify central objects to the story.

Results show our method’s advantage; the most influential objects need not be the most frequent.