Category-specific video summarization

Danila Potapov Matthijs Douze Zaid Harchaoui Cordelia Schmid
Inria Grenoble

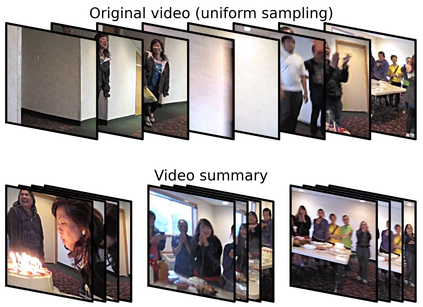
Introduction

A *video summary*

- built from subset of temporal segments of original video
- conveys the most important details of the video

Our approach

- produce *visually coherent* temporal segments
  - no shot boundaries, camera shake, etc. inside segments
- identify important parts
  - *category-specific importance*: a measure of relevance to the type of event



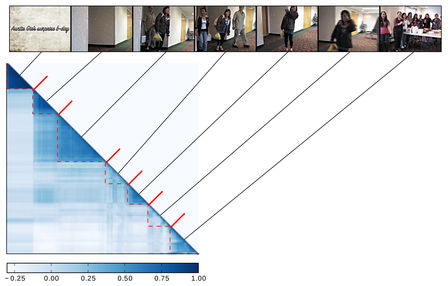
Original video, and its video summary for the category "Birthday party".

Contributions

- temporal video segmentation algorithm
- novel approach for supervised video summarization
- *MED-Summaries*: dataset for evaluation of video summarization

Kernel Temporal Segmentation

- input: robust frame descriptor (SIFT + Fisher Vector)
- kernelized Multiple Change-Point Detection algorithm
- solved exactly with dynamic programming in $O(mn^2)$
- optimization criterion: minimize the sum of within-segment variances
- automatic calibration of the number of change points with a BIC-like regularizer

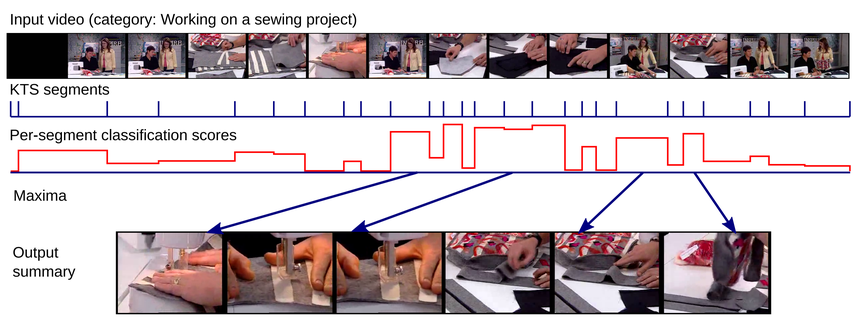


Kernel matrix and temporal segmentation of a video

Supervised summarization

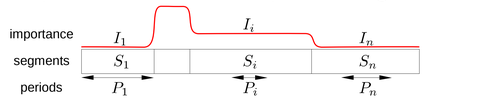
- **Training:** Train a linear SVM from a set of videos with just video-level class labels.
- **Testing:** Score segment descriptors with the classifiers trained on full videos. Build a summary by concatenating the most important segments of the video.

Overview



MED-Summaries dataset

- 100 test videos (= 4 hours) from Trecvid MED 2011
- multiple annotators
- 2 annotation tasks:
  - segment boundaries (median duration: 3.5 sec.)
  - segment importance (grades from 0 to 3)
- additional *period* attribute for repetitive segments



Central frame for each segment with importance annotation for category "Changing a vehicle tyre".

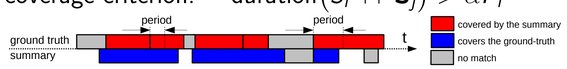
lear.inrialpes.fr/people/potapov/med_summaries.php

Evaluation metrics for summarization

- often based on user studies
  - time-consuming, costly and hard to reproduce

Our approach

- ground truth segments $\{S_i\}_{i=1}^{m}$
- computed summary $\{\widetilde{\mathbf{S}}_j\}_{j=1}^{\tilde{m}}$
- coverage criterion: $\text{duration}\left(S_i \cap \widetilde{\mathbf{S}}_j\right) > \alpha P_i$



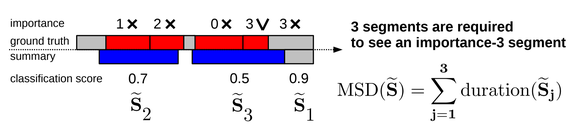
- *importance ratio* for summary $\widetilde{\mathbf{S}}$ of duration $\mathbf{T}$

$$\mathcal{I}^*(\widetilde{\mathbf{S}}) = \frac{\mathcal{I}(\widetilde{\mathbf{S}})}{\mathcal{I}^{\max}(\mathbf{T})}$$

total importance covered by the summary

max. possible total importance for a summary of duration $\mathbf{T}$

- a *meaningful summary* covers a ground-truth segment of importance 3



$$\text{MSD}(\widetilde{\mathbf{S}}) = \sum_{\mathbf{j}=\mathbf{1}}^{\mathbf{3}} \text{duration}(\widetilde{\mathbf{S}}_\mathbf{j})$$

Meaningful summary duration (MSD): minimum length for a meaningful summary

- segmentation *f-score*: match when overlap/union $> \beta$

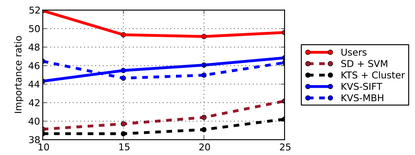
Baselines

- **Users**: keep 1 user in turn as a ground truth for evaluation of the others
- **SD + SVM**: shot detector (Massoudi, 2006) for segmentation + same importance scoring
- **KTS + Cluster**: same segmentation + k-means clustering for summarization
  - sort segments by increasing distance to centroid

Results

Method	Segmentation Avg. f-score (higher better)	Summarization Med. MSD (s) (lower better)
Users	49.1	10.6
SD + SVM	30.9	16.7
KTS + Cluster	**41.0**	13.8
KVS	**41.0**	**12.5**

Segmentation and summarization performance



Importance ratio for different summary durations.

Example summaries