Automatic summarization of video data

Presented by Danila Potapov

Joint work with:

Matthijs Douze   Zaid Harchaoui   Cordelia Schmid
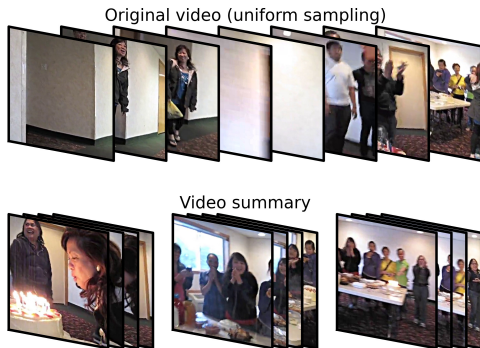
LEAR team, Inria Grenoble

Khronos-Persyvact Spring School   1.04.2015

Definition

A *video summary*

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

Original video (uniform sampling)

Video summary

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

Overview of 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


Input video (category: Working on a sewing project)

KTS segments

Per-segment classification scores

Maxima

Output summary
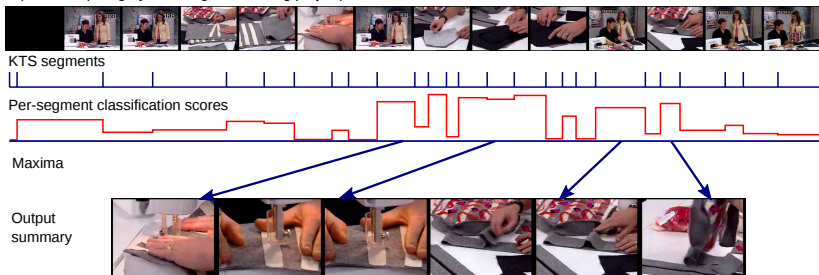
Contributions

- 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
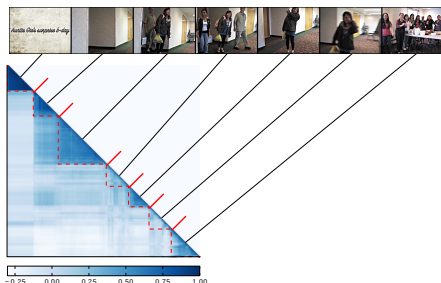
Kernel matrix and temporal segmentation of a video

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.

Input video (category: Working on a sewing project)

KTS segments

Per-segment classification scores

Maxima

Output summary
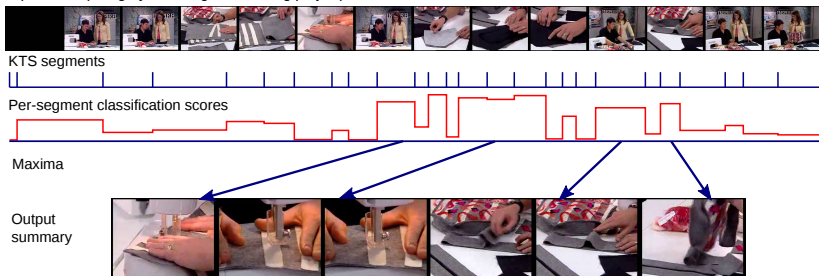
MED-Summaries dataset

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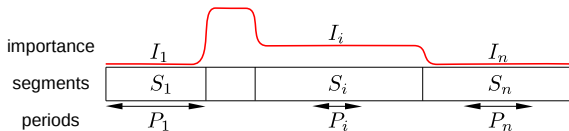
Central frame for each segment with importance annotation for category "Changing a vehicle tyre".

Evaluation metrics for summarization (1)

- ▶ often based on user studies
  - ▶ time-consuming, costly and hard to reproduce
- ▶ **Our approach:** rely on the annotation of test videos
- ▶ ground truth segments $\{S_i\}_{i=1}^{m}$
- ▶ computed summary $\{\widetilde{\mathbf{S}}_j\}_{j=1}^{\tilde{m}}$
- ▶ coverage criterion: $\mathrm{duration}\big(S_i \cap \widetilde{\mathbf{S}}_j\big) > \alpha P_i$
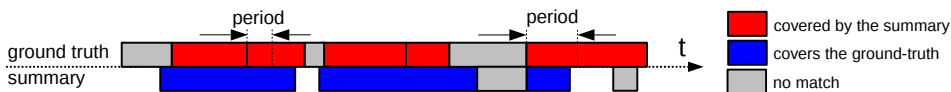
- ▶ *importance ratio* for summary $\widetilde{\mathbf{S}}$ of duration **T**

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

total importance
covered by the summary

max. possible total importance
for a summary of duration **T**

Evaluation metrics for summarization (2)

- a *meaningful summary* covers a ground-truth segment of importance 3
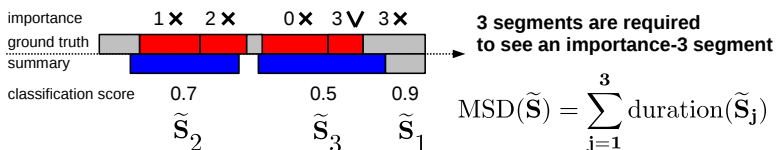
Meaningful summary duration (MSD): minimum length for a meaningful summary

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

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

Experiments

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
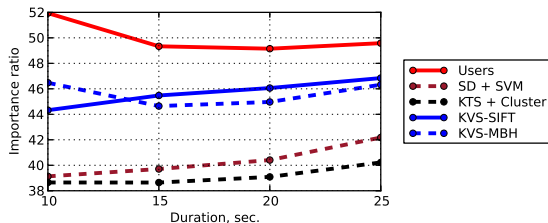 - sort segments by increasing distance to centroid

Our approach

- **KVS** = **KTS** + **SVM**

Results

Method	Segmentation Avg. f-score	Summarization Med. MSD (s)
	higher better	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

Examples summaries

Birthday party

Uniform sampling

Our video summary

0.189   0.151   0.122   0.077   0.055

Changing a vehicle tire

Uniform sampling

Our video summary

0.096   0.081   0.036   0.034   0.026

Parade

Uniform sampling

Our video summary

0.309   0.089   0.064   0.047   0.032

Conclusion

- KVS delivers short and highly-informative summaries, with the most important segments for a given category

- KVS is trained in a semi-supervised way
  - does not require segment annotations in the training set

- MED-Summaries — publicly available dataset
  - annotations and evaluation code available online:
    http://lear.inrialpes.fr/people/potapov/

Thank you for your attention!

References

- MED-Summaries dataset `lear.inrialpes.fr/people/potapov/med_summaries.php`
- D. Potapov, M. Douze, Z. Harchaoui, C. Schmid "Category-specific video summarization", ECCV 2014

- **Related work**
    - M. Sun et al. "Ranking Domain-specific Highlights by Analyzing Edited Videos", ECCV 2014
    - M. Gygli et al. "Creating Summaries from User Videos", ECCV 2014