Category-specific video summarization
Supplementary material

Paper ID 1153

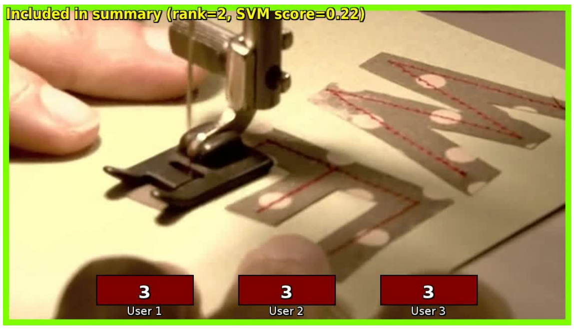
This supplement gives more details on several aspects on the proposed approach **KVS**. First, in Sec. A, we present an example of video summary output by our approach. This is a companion section to the attached video (file `demo1153.avi`). Then, in Sec. B, we give a description of the categories in the **MED-Summaries** dataset. In Sec. C, we describe our annotation interface. Finally, in Sec. D, we provide a convergence proof for the Kernel Temporal Segmentation algorithm.

A. Demo of a video summary

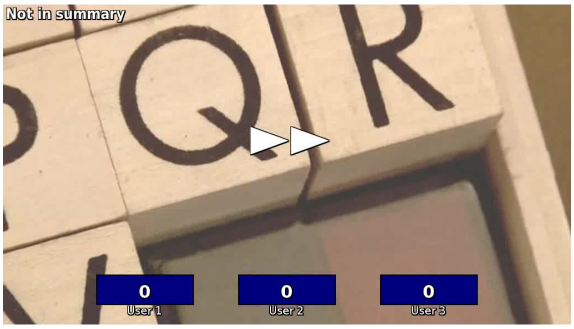
The video clip (file `demo1153.avi`) shows the proposed **KVS** video summarization approach in action. We show examples of video summaries for 2 event categories: “Birthday party” and “Working on a sewing project”. A video summary temporally assembles the sequence of segments with the highest importance scores. The duration of the summaries presented here is limited to 15 seconds.

Each video shows both the segments that are included in the summary and the segments that are not included. A green frame lights up for segments that are included in the summary (see left image (a) of Fig. 6). Video segments that are not included in the summary are shown with a fast-forward (see right image (b) of Fig. 6).

Each of the videos was annotated by 3 users. We display the ground-truth importance scores assigned by each user as color bars with numbers from 0 to 3. Figure 6 shows two frames from a video summary.



(a) Included in summary.



(b) Not in summary.

Figure 6: Snapshots of a video summary corresponding to the “Working on a sewing project” category. Left part shows a frame from a video segment that is **included in the summary**, with a green frame and the label “Included in the summary” on the top-left. Importance scores assigned by the 3 users appear in the bottom of the image. Right part shows a frame from a video segment that is **not included in the summary**, without any colored frame and with the label “Not included in the summary” on the top-left. Again, importance scores appear in the bottom.

B. Category description

Trecvid 2011 Multimedia Event Detection [1] dataset provides textual descriptions for each event category. The descriptions were shown to users as a reference for importance annotation. As an example the "Birthday party" category is defined as follows:

> *Event name*: Birthday party
> *Definition*: An individual celebrates a birthday with other people
> *Explication*: A birthday in this context is the anniversary of a person's birth. Less commonly, the term "birthday" can be used to refer to the anniversary of an organization's establishment, but a celebration for an organization does not satisfy the event definition.
> A birthday celebration is a gathering of people who have been invited by the host or hosts to come to a set location (often a private home, sometimes a restaurant, bar, nightclub, park, or other public venue) to socialize in honor of the person(s) whose birthday it is (the birthday celebrant(s)).
> Birthday parties, as with other parties/celebrations, will typically feature an assortment of food and beverages. Birthday parties are often accompanied by colorful decorations, such as balloons and streamers, and some people may wear cone-shaped "birthday hats". The decorations may include signs or banners displaying a message for the birthday celebrant. Often, especially for children's parties, guests will bring cards and/or gifts wrapped in shiny/colorful paper or bags, which will be opened by the birthday celebrant(s), or by their parents/siblings if the birthday celebrants are too young to open the gifts by themselves. A cake (or sometimes cupcake or other food item) with lit candles, called the "birthday cake", is often served. The song "Happy Birthday to You" may sung by the guests while the birthday cake with lit candles is carried to a table or counter where the birthday celebrant(s) are seated. The birthday celebrant(s) then blow out the candles, usually after the song is finished, and the guests then clap and cheer. Birthday parties may also involve games or other organized group activities.
> *Evidential description*:
> *scene*: indoors (a home, a restaurant) or outdoors (backyard, park); day or night
> *objects/people*: decorations (balloons, streamers, conical hats, etc), birthday cake (often with candles), birthday celebrant, guests, gifts
> *activities*: singing, blowing out candles on cake, playing games, eating, opening gifts
> *audio*: singing "Happy Birthday to You"; saying happy birthday; laughing; sounds of games being played

C. Annotation interface

In order to get ground-truth annotation of temporal segments and importance, we developed a web-based interface. Figure 7 shows a screenshot at the point when the annotation is finished.

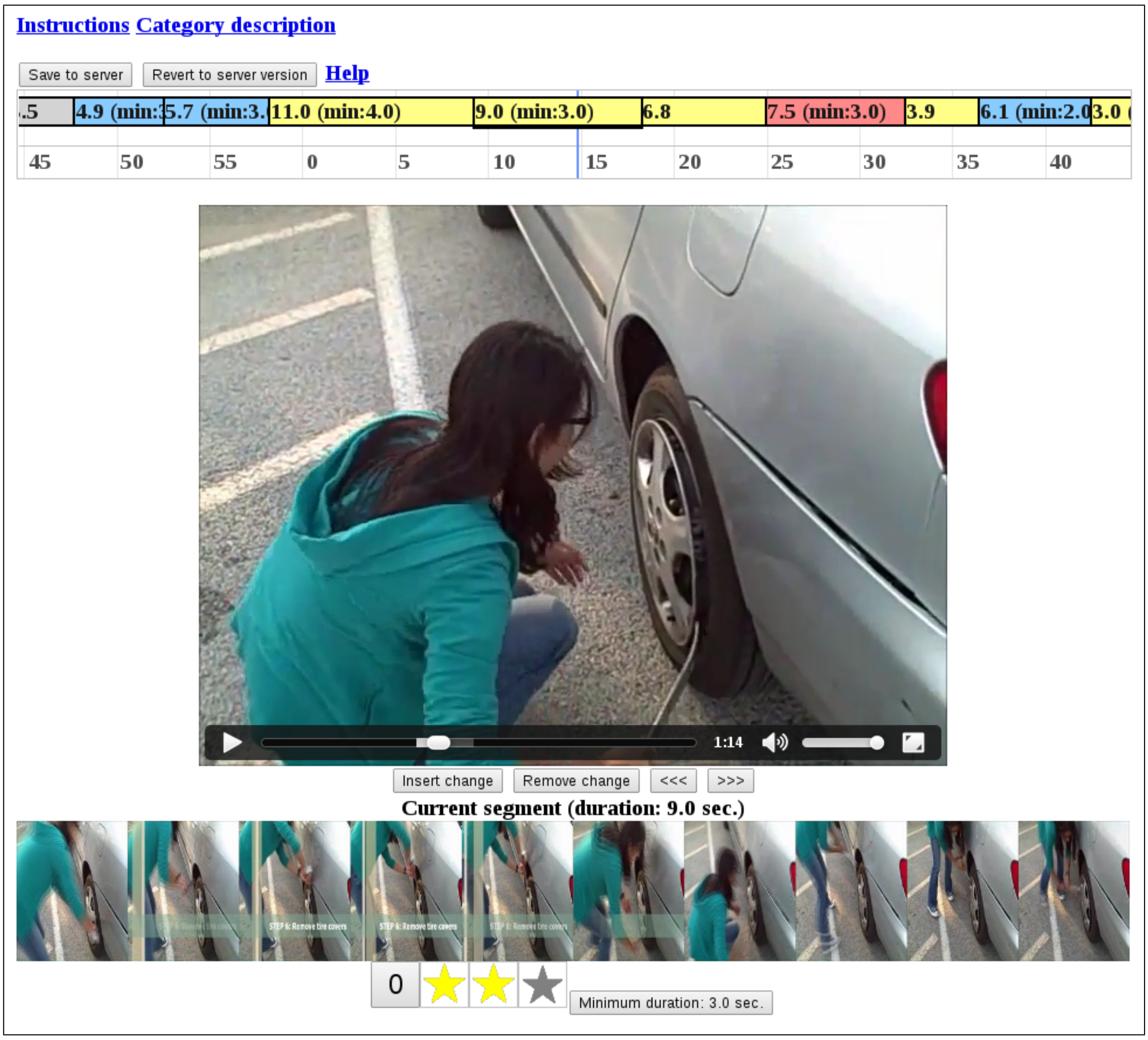


Figure 7: Interface for annotation of temporal segments and importance. Video from "Changing a vehicle tire" category.

D. Properties of Algorithm 1

We refer to Algorithm 1 for Kernel Temporal Segmentation as A_1 for brevity. A_1 has the following properties:

1. **(Optimality)** Let $\theta^* = \{m^*; t_0, \ldots, t_{m^*-1}\}$ be a solution returned by A_1. Optimization objective (1) attains in θ^* the exact global optimum. [1]

2. **(Termination)** A_1 stops after $O(n^2)$ operations.

Proof

(Optimality)
To disambiguate the notation, we will refer to $L_{m,n}$ as defined in Step 3 of A_1, not in formula (2). Let us define for $j = 1, \ldots, n$ (we assume $t_{-1} = 0$):

$$f_j(t_0, \ldots, t_{m-1}) := \sum_{i=0}^{m-1} v_{t_{i-1},t_i} + v_{t_{m-1},j} \tag{7}$$

$$f_j^*(m) := \min_{\substack{t_0,\ldots,t_{m-1} \\ 0<t_0<\ldots<t_{m-1}<j}} f_j(t_0, \ldots, t_{m-1}) \tag{8}$$

Then the objective (1) writes as:

$$\min_{\substack{m;\ t_0,\ldots,t_{m-1} \\ 0<t_0<\ldots<t_{m-1}<n}} f_n(t_0, \ldots, t_{m-1}) + Cg(m, n) \tag{9}$$

To prove the optimality property we will show that Step 3 of A_1 computes the function $f_j^*(i)$.
From Definitions (7)-(8): $\forall j = 1, \ldots, n \quad \forall m = 1, \ldots, m_{\text{max}}$ s.t. $m < j$:

$$f_j^*(m) = \min_{m-1<t_{m-1}<j} \min_{\substack{t_0,\ldots,t_{m-2} \\ 0<t_0<\ldots<t_{m-1}}} f_j(t_0, \ldots, t_{m-1}) \tag{10}$$

$$= \min_{m-1<t_{m-1}<j} \min_{\substack{t_0,\ldots,t_{m-2} \\ 0<t_0<\ldots<t_{m-1}}} \sum_{i=0}^{m-2} v_{t_{i-1},t_i} + v_{t_{m-2},t_{m-1}} + v_{t_{m-1},j} \tag{11}$$

$$= \min_{m-1<t<j} \min_{\substack{t_0,\ldots,t_{m-2} \\ 0<t_0<\ldots<t}} \sum_{i=0}^{m-2} v_{t_{i-1},t_i} + v_{t_{m-2},t} + v_{t,j} \tag{12}$$

and we finally get

$$f_j^*(m) = \min_{m-1<t<j} f_t^*(m-1) + v_{t,j}\ . \tag{13}$$

We also note that by definition $f_j^*(0) = v_{0,j}$. Therefore we state that Step 3 of A_1 computes $L_{i,j} = f_j^*(i)$, because reccurrent functions $L_{i,j}$ and $f_j^*(i)$ are the identical, which proves the optimality property for the fixed m.

At Step 4, A_1 selects the optimal m^*, because the penalty term $Cg(m, n)$ in (9) does not depend on change-point positions. Finally, the optimal change-points are identified for the fixed value m^* at Step 5.

(Termination)
The following table shows computational complexities of each step of the algorithm

Step	Complexity	Commentary
1	$O(n^2)$	—
2	$O(n^2)$	Cumulative sums are precomputed as in [2]
3	$O(m_{\text{max}} n^2)$	—
4	$O(m_{\text{max}})$	—
5	$O(m_{\text{max}})$	Optimal values are stored at Step 2

We assume the constant m_{max} to be fixed. Then the algorithm stops after $O(n^2)$ operations.

[1] Equation numeration starts in the main paper.

References

[1] Over, P., Awad, G., Fiscus, J., Antonishek, B., Michel, M., Smeaton, A.F., Kraaij, W., Quénot, G., et al.: An overview of the goals, tasks, data, evaluation mechanisms and metrics. In: Proceedings of TRECVID. (2011)

[2] Crow, F.C.: Summed-area tables for texture mapping. In: ACM SIGGRAPH Computer Graphics. Volume 18. (1984) 207–212