Supplementary file for:

Story-Driven Summarization for Egocentric Video, Paper ID 1257

This document contains:

- More details on the search procedure in Sec. 3.3 of the submission.
- User interface for our human subject studies, in which human subjects compared our results to each of the baselines.
- User interface for obtaining ground truth important objects from a given video though MTurk.

Attached mpg video:

- A single video containing examples of video summarization results (accelerated): We show two example summaries each for the UTE and ADL datasets. For each example, we first display our method's summary, followed by that of the uniform sampling, shortest path, and object-driven [14] baselines. Object-driven [14] is not available on the ADL data as explained in the main text. Please review the descriptions within the video to help interpret the results.

1. Algorithm to search for the optimal chain of subshots (Sec. 3.3)

Algorithm 1 findGoodChain($\mathcal{V}$,k)

$s = 1$
$t =$ last node of $\mathcal{V}$
for each $(u, v) \in E$ **do**
  $val_{(u,v)} = \lambda_s S([u, v]) + \lambda_i I([u, v]) + \lambda_d D([u, v])$
end for
$pq =$ new priority queue
pq.insert($\langle [s], 0 \rangle_{exact}$)
while pq is not empty **do**
  $(\pi, val_\pi) =$ pq.top
  if π is marked 'approx' **then**
    $newval_\pi = \lambda_s S(\pi) + \lambda_i I(\pi) + \lambda_d D(\pi)$
    pq.insert($\langle \pi, newval_\pi \rangle_{exact}$))
  else
    $u = \pi$.lastNode
    if $t - u < \varepsilon_1$ **then**
      return π
    end if
    for $v < u + \varepsilon_2$ **do**
      $newval_{[\pi,v]} = \lambda_i I([\pi, v]) + \lambda_d D([\pi, v])$
      $+ \min(val_\pi - \lambda_i I(\pi) - \lambda_d D(\pi), \lambda_s S([u, v])$
      pq.insert($\langle [\pi, v], newval_{[\pi,v]} \rangle_{approx}$)
    end for
  end if
end while

The above shows the detailed algorithm for searching for the optimal chain of subshots, as described in Sec. 3.3 of the submission. The algorithm is similar to one in CTD [24], but expands the objective function to incorporate our importance and diversity terms. Note, the threshold ε_1 is used to allow some flexibility for the end point of the selected chain. The threshold ε_2 is used to control how far two neighboring can be apart from each other. These two parameters are used for efficiency only, and are set to $0.5\times(t-s+1)/(k-1)$ and $1.5\times(t-s+1)/(k-1)$ in our results. See our submission for the definition of all other variables/functions appearing above.

2. User interface for our user studies

Instructions:

- We are showing a **visual summary of a video** captured from an ego-centric (first-person) camera. The **GOAL** of the visual summary is to capture the ***main*** **story** in the original video and also show the ***progress*** **of the story**. In particular, a **good summary** should make it clear how one sub-event ***LEADS TO*** the next.
  - For example, let's consider three sub-events A, B, and C. Sub-event A is about walking with a book in the library. Sub-event B is about borrowing the book at the check-out counter. Sub-event C is about looking at somebody in the library. **It is clearer that sub-event A leads to sub-event B compared to sub-event A and C.**
- The visual summaries consist of several short video clips. We display the summary as the series of clips separated by arrows.
- Your task is to compare two variants of the summary, and decide which better reflects the story, as follows:
  1. Click the play button to watch the speed-up version of the original video.
  2. Answer question 1 by writing the text description of the story in the video.
  3. Click the play button to watch each video clip (each is a few seconds).
  4. Answer question 2 by selecting set 1 or set 2.
  5. Indicate your confidence level.
  6. Click "Submit" to submit your answer.

First, please watch the speed-up version of the original video.

Question 1: What is the story in this video? Please write down the text decription in a few sentences. Be sure to note that series of sub-events flow from the beginning to the end.

Answer:

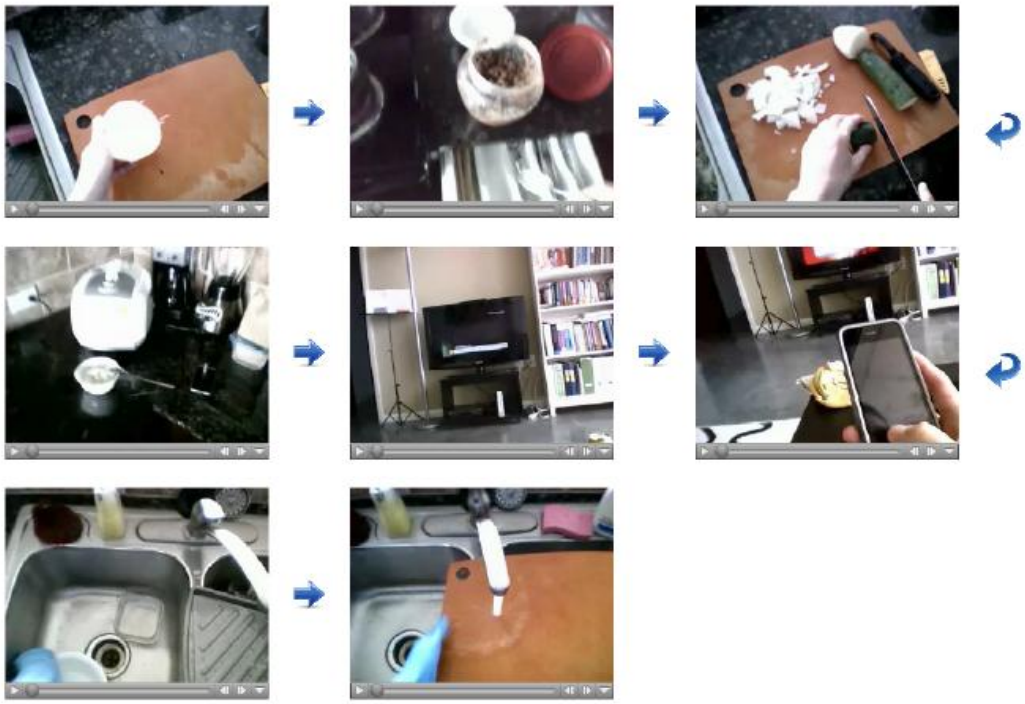
Question 2: Which summary shows better progress of the story? In particular, for which case (set 1 or set 2) is it more apparent that

A. **the earlier sub-events *lead to* later ones** (Please pay EQUAL attention to ALL the progresses of sub-events. See the example in the instruction to understand this point better),
B. **redundancy is avoided, and**
C. **each important sub-event from the original video is represented well.**

Feel free to watch clips multiple times, as needed, to decide on your answer.

Answer:

SET 1



SET 2

In 1-2 sentences, please describe in the textbox below the ^reason^ for your selection of set 1 or set 2. Then rate your confidence.

Confidence level of the answer:

- I am almost guessing which is better.
- It is not obvious, but the one I chose seems slightly better.
- The one I chose is a bit better.
- The one I chose is clearly better.

3. User interface for obtaining important objects

Instructions:

- We are showing **a short video** captured from an ego-centric (first-person) camera.
- Some objects that appear in the video are important to the main storyline, while others are unimportant. An important object is one that is essential to following the main sub-events.For example, things used by the camera wearer are likely important. Things at the place where the camera wearer is looking at are likely important.
- Your task is to decide which ^**objects**^ appearing in the video are **important** to the story, as follows:
  1. Click the play button to watch the speed-up version of the video.
  2. Answer question 1 by writing the text description of the story in the video.
  3. Answer question 2 by selecting "very important", "somewhat important", or "not important" for each object listed below, according to your understanding of the video.
  4. Click "Submit" to submit your answer. Next page will show your confirmation code. Enter it in the MTurk answer box to complete the job.

First, please watch the speed-up version of the video. Please pay attention to the ^**important objects**^ appeared in the video.

Question 1: What is the story in this video? Please write down the text decription in a few sentences. Be sure to note the series of sub-events that flow from the beginning to the end.

Answer:

Question 2: Which objects listed in the following are important to the original video? Note that all these objects appeared in the video. Feel free to watch the video or navigate forward and backward multiple times, as needed, to decide on your answer.

Answer:

milk:	very important	somewhat important	not important
condiments:	very important	somewhat important	not important
electric keys:	very important	somewhat important	not important
fridge:	very important	somewhat important	not important
knife/spoon/fork:	very important	somewhat important	not important
microwave:	very important	somewhat important	not important
mug/cup:	very important	somewhat important	not important
oven/stove:	very important	somewhat important	not important
pan/pot:	very important	somewhat important	not important
tap:	very important	somewhat important	not important
towel paper:	very important	somewhat important	not important
trash can:	very important	somewhat important	not important
tv:	very important	somewhat important	not important