Supplementary file for:

Discovering Important People and Objects for Egocentric Video Summarization, CVPR 2012

Yong Jae Lee, Joydeep Ghosh, and Kristen Grauman

This document:

- Amazon Mechanical Turk interfaces described in Sec. 3.2 of the main paper.
- Qualitative examples of regions selected to be important by (1) our method, (2) the object-like method of [3], (3) the object-like method of [6], and (4) the bottom-up saliency method of [30].

Attached mpg videos:

- Summarization results in the form of videos: Our summaries are named "userX-ours.mpg", while the uniform sampling baseline summaries are named "userX-baseline.mpg". For our results, the frame boundary colors indicate different events. Some faces are blocked out to preserve anonymity. The summaries are among the various we tested and quantified results for in Fig. 7 and Table 1 of the main paper.
- A clip from a sample raw video: "sample.mpg".

(1) Mechanical Turk interfaces

Interface 1:

1. First, watch the video. **The video is 3 minutes long.** (If the video does not show, try refreshing the webpage. The video can take a few seconds to load, depending on your internet connection.)

2. Then, describe every ***visible*** **important object/person** in the video. There will be 1 to 5 important objects/people. These are key items/players that are essential to the "story"; i.e., things that would be necessary to create a summary of the video. For example, **objects/people that frequently appear, objects/people that the camera wearer interacts with**, are some things that could be considered important. Be as descriptive as possible (e.g., *(1) The coffee mug next to the stack of napkins. (2) The spaghetti that the camera wearer is eating*).
The "submit" button will be enabled after you complete the task.

You MUST see these examples before you go on!

We will be checking results and may reject jobs or ban users that do not follow the above instructions.

(1)
(2)
(3)
(4)
(5)

Interface 2:

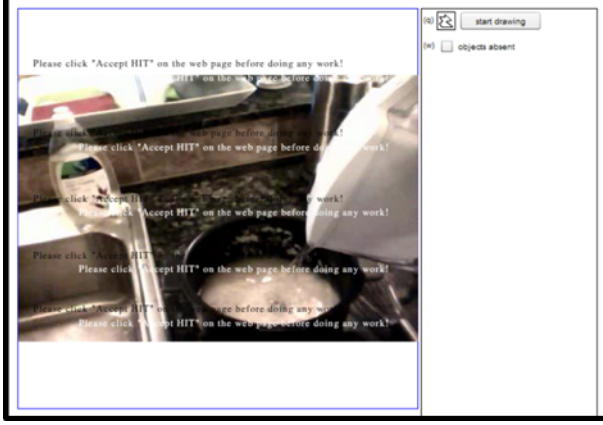
1. First, **read the text descriptions** of the objects to find. These are the objects you are looking for in the images!

2. Second, find those objects in each of the six images below. If the image **does not contain any of the described objects**, check the button "*objects absent*". Otherwise, click "*start drawing*" to draw a **tight-fitting boundary of one described object** in the image.* Click anywhere on the object's boundary to start the drawing, and click on another point to extend the drawing. You must close the boundary by clicking back on the first point. You can click "*undo*" to remove the last drawn boundary segment, and "*cancel*" to start over. **If the image contains multiple described objects, select the most central, prominent one to annotate.**
If the text describes a person whose face is visible, then draw a boundary around only the face.

Please look at these examples and the demo video on the right before you go on!

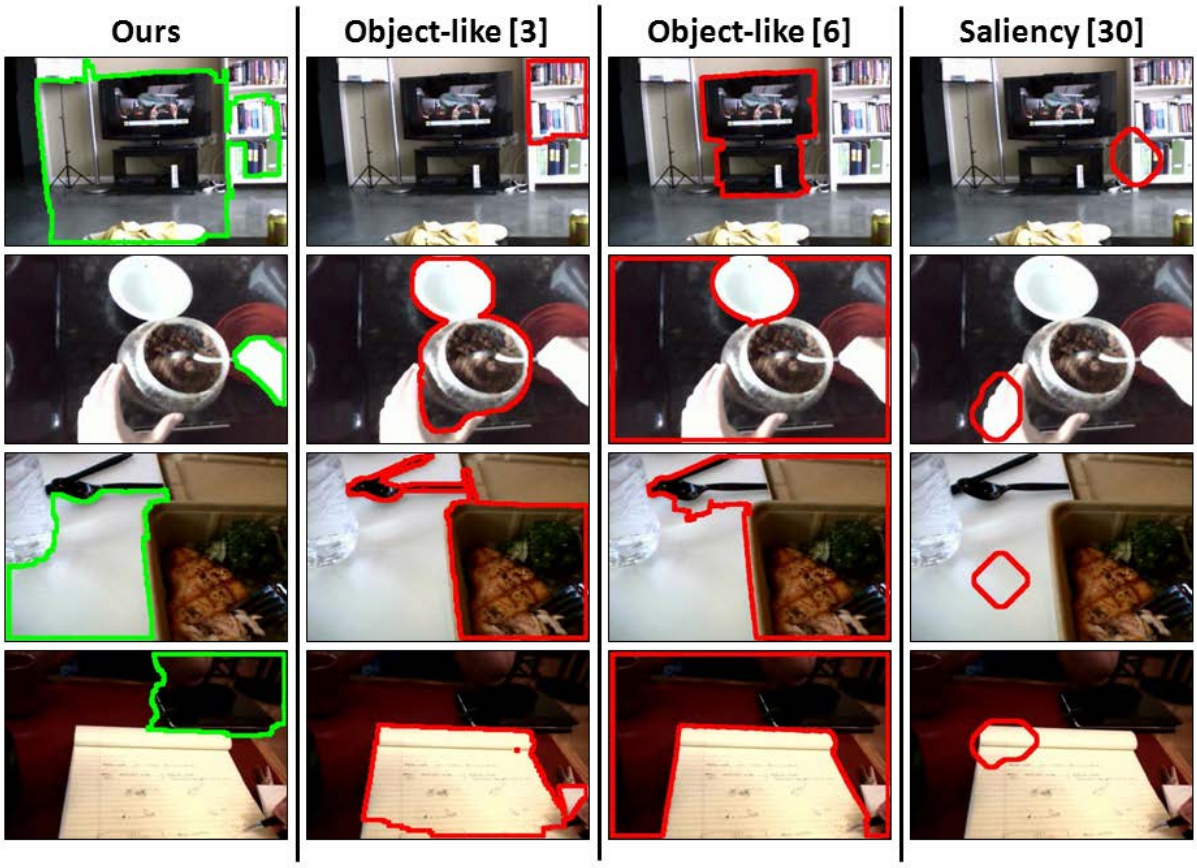
You must click "submit results" on all six images to complete the hit.

Objects to find: ***(1) black pot with white rice in water (2) white rice cooker (3) television (4) man in glasses wearing orange t-shirt***



(2) Important region examples

This figure shows examples of correct predictions made by our method. The high-level saliency detection methods [3,6] focus on detecting any prominent object. Therefore, unlike our method, they are unable to distinguish those that may be important to a camera wearer. The low-level saliency detection method [30] fails to find object-like regions, and instead produces local estimates of saliency. See Fig. 5 in the main paper for quantitative results on this part of the method. **Best viewed in color.**



This figure shows examples of incorrect predictions made by our method. The high-level saliency detection methods [3,6] produce better predictions for these examples. In the first example, our method produces an under-segmentation of the important object and includes regions surrounding the television. In the second example, our method incorrectly detects the user's hand to be important, while in the third and fourth examples, it determines background regions to be important due to their high frequency. See Fig. 5 in the main paper for quantitative results on this part of the method. **Best viewed in color.**