Making 360° Video Watchable in 2D: Learning Videography for Click Free Viewing

Yu-Chuan Su Kristen Grauman
The University of Texas at Austin

The supplementary materials consist of:

A Details for the HumanEdit-based metric

B Complete computational cost experiment results

C Annotation interface introduction

Please refer to the project webpage for the annotation interface demonstration and example output videos.

A. HumanEdit-Based Metric

We deploy two different strategies for pooling the frame-wise **overlap**:

- **Trajectory pooling** rewards algorithm outputs that are similar to *at least one* HumanEdit trajectory over the entire video. It first computes the per video overlap between each algorithm-generated trajectory and HumanEdit trajectory using the average per frame overlap. Each algorithm output is then assigned the score as the overlap with its most similar HumanEdit trajectory.

- **Frame pooling** rewards algorithm outputs that are similar to *any* HumanEdit trajectory at each frame. For each algorithm output, it first scores each frame using its overlap to the most similar HumanEdit trajectory. The output trajectory is then assigned the score as the average per frame score.

B. Computational Cost Results

For completeness, Fig. 1 shows the computational cost versus output quality for all metrics as noted in footenote 6 of the main paper. Due to space limits, the main paper includes the same result for Distinguishability and Trajectory overlap in Fig 8. OURS W/ FAST significantly outperforms AUTOCAM [1] in all metrics. It performs similarly to OURS in Transferability and Frame Overlap but worse in the HumanCam-Likeness metric. This is consistent with the Distinguishability metric and is possibly due to the distortion in 104.3° FOV. Note the HumanCam-Likeness metric

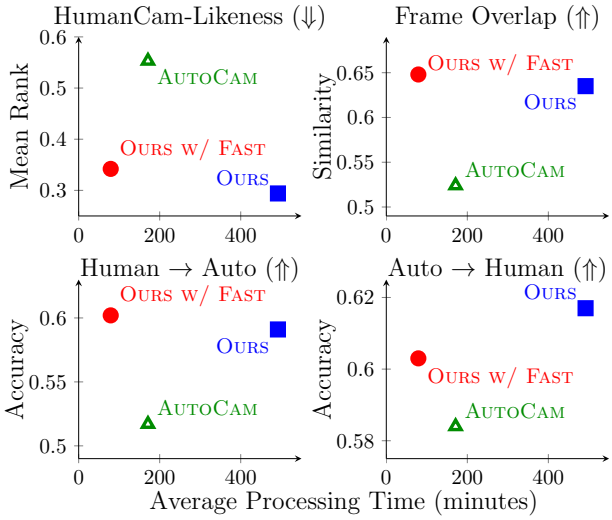


Figure 1: Computational cost versus output quality. The arrows in title indicate higher scores better (⇑) or lower scores better (⇓). The results are consistent with the distinguishability and trajectory overlap metrics in the main paper, and were pushed to supp. due to space limits.

is measured by the normalized ranking and is a relative metric, so the absolute value depends on the number of methods evaluated and is different from the results in the paper.

C. HumanEdit Interface

The annotation interface displays the 360° video in equirectangular projection so the editors can see all the visual content at once. The interface also extends the panoramic strip by 90° on both sides to mitigate problems due to discontinuities at the edge. See Fig. 2. The editors are instructed to move the cursor to direct a virtual NFOV camera, where the frame boundaries are backprojected onto the panoramic strip in real time to help the editors see the content they capture. The editors can also control the focal length of the virtual camera. The available focal lengths are the same as those available to the algorithm, and the inter-

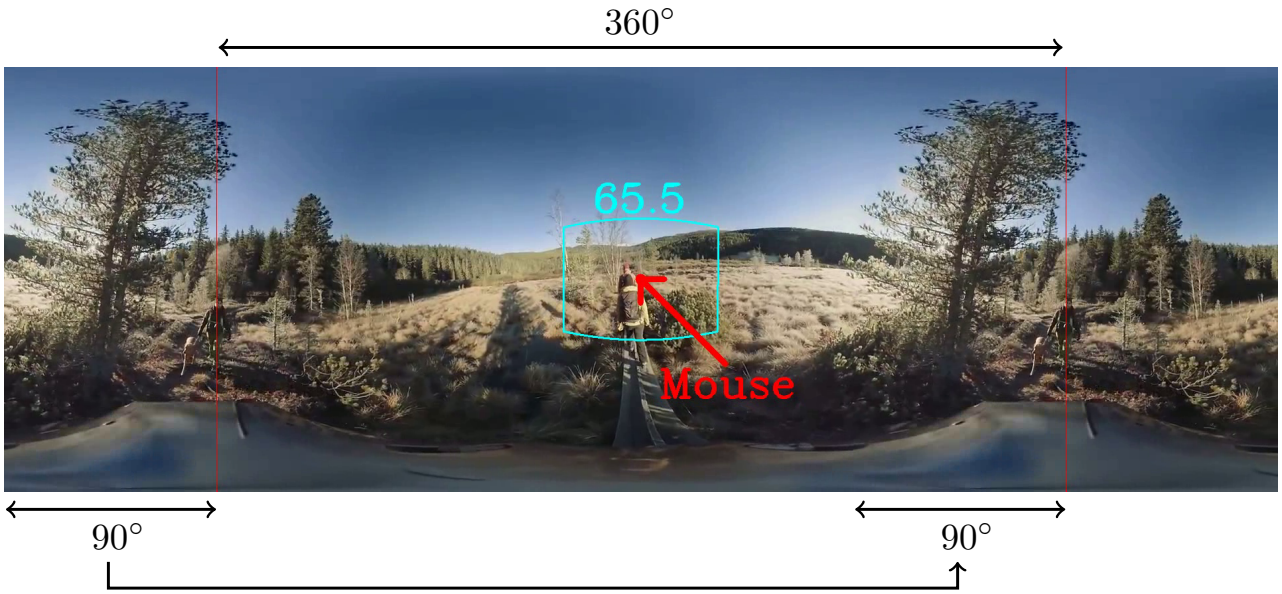


Figure 2: The interface shows the video in panoramic strip. It further expands both side by $90°$.

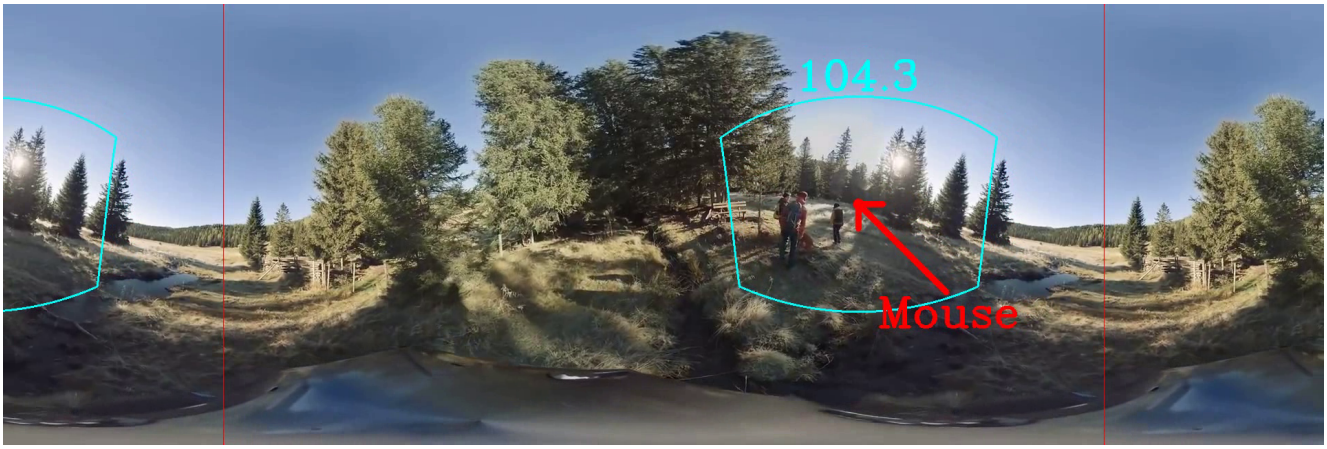
face will switch to the next available focal length when the editor presses the button for zoom in/out. See Fig. 3.

For each $360°$ video, we ask the editors to watch the full video in equirectangular projection first to familiarize themselves with the content. Next, we ask them to annotate *four* camera trajectories per video. For each of the four passes, we pan the panoramic strip by the angle of $[0°, 180°, 0°, 180°]$ to force the editors to consider the trajectories from different points of view. Finally, for the first two trajectories of the first two videos annotated by each editor, we render and show the output video to the editor right after the annotation to help him understand what the resulting video will look like.

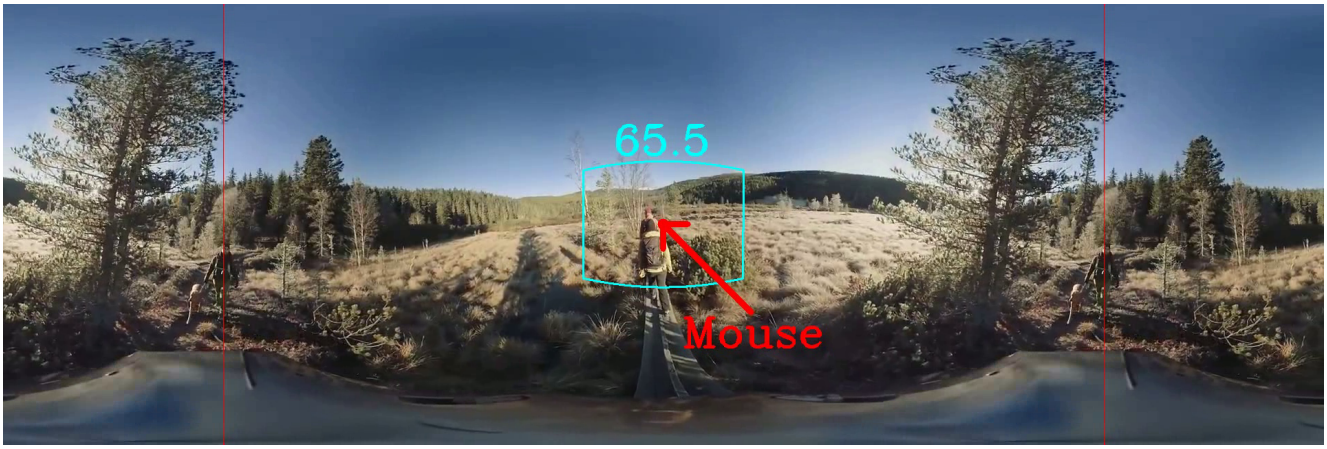
Also see our project webpage for video examples of the interface in action.

References

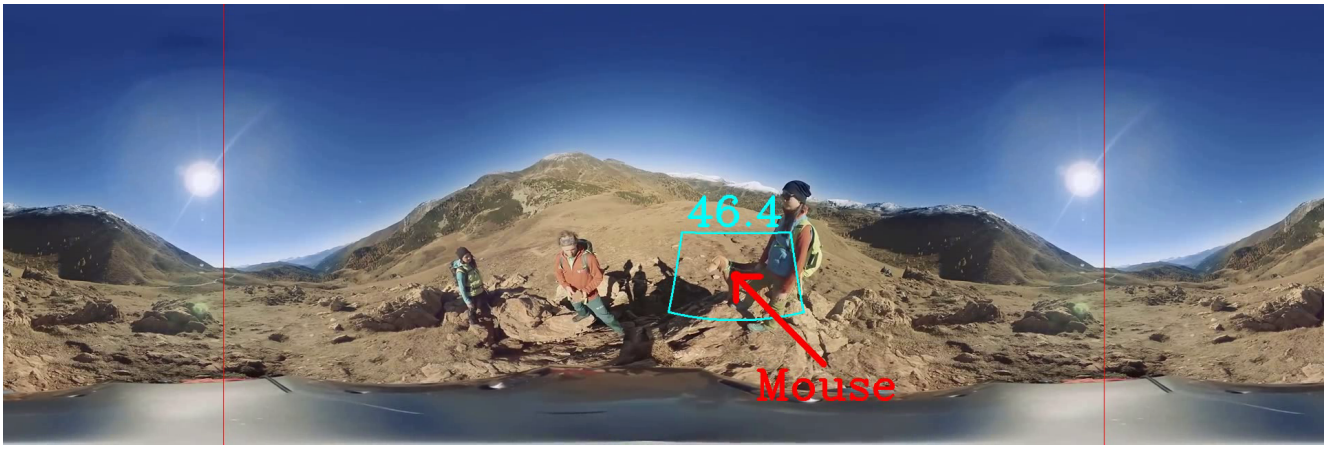
[1] Y.-C. Su, D. Jayaraman, and K. Grauman. Pano2vid: Automatic cinematography for watching $360°$ videos. In *ACCV*, 2016. 1



(a) Zoom out.



(b) Original.



(c) Zoom in.

Figure 3: The interface allows the human editors to control the FOV.