Pano2Vid: Automatic Cinematography for Watching 360° Video

Yu-Chuan Su, Dinesh Jayaraman, and Kristen Grauman
The University of Texas at Austin
http://vision.cs.utexas.edu/projects/Pano2Vid/

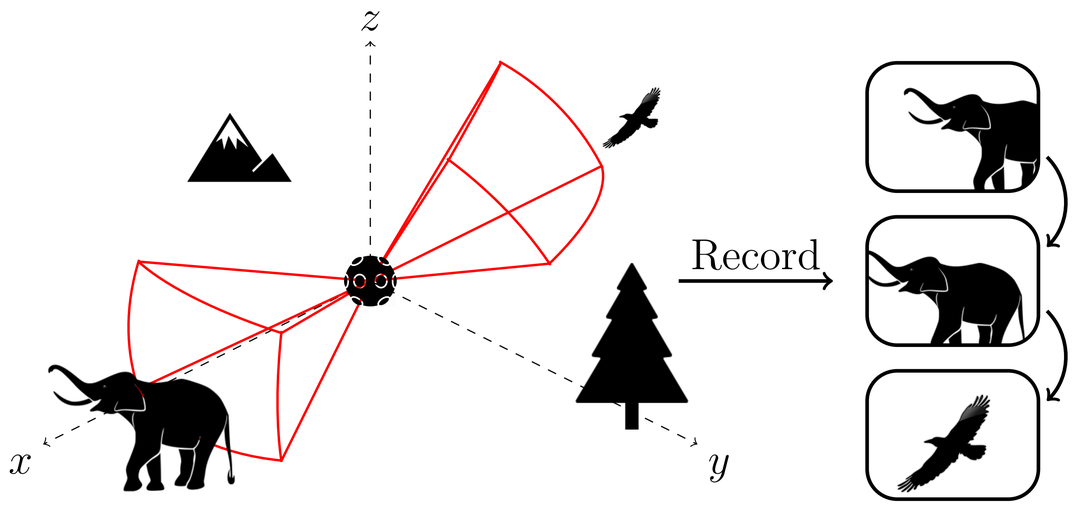
1. Pano2Vid

360° Camera

Challenge for watching 360° video:
How to find the right direction to watch?

Pano2Vid Definition

Input: 360° video
Output: "natural-looking" normal-field-of-view (NFOV) video
Task: **control** the virtual camera **direction**



2. Proposed Solution – AutoCam

Key Idea: Learn videography tendencies from Web videos

- Learn diverse capture-worthy content
- Pickup proper composition

Spatio-Temporal Glimpses

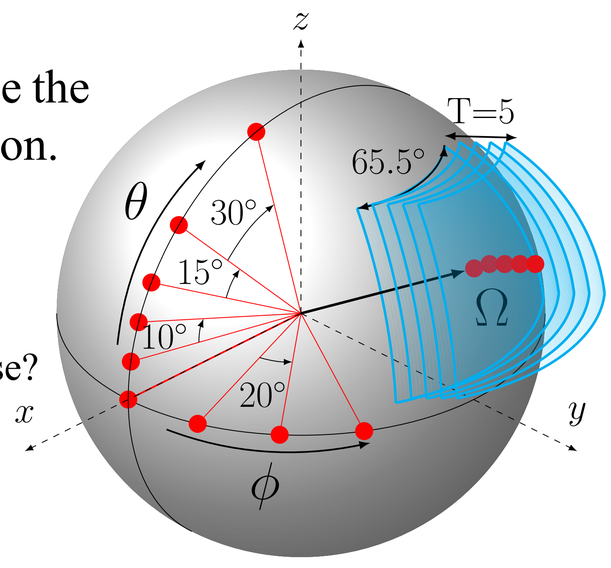
- Short NFOV video extracted from 360° video
- Makes 360° content comparable with NFOV videos
- Fixed camera parameter & direction

Sample ST-glimpses

Sample ST-glimpses and reduce the problem to ST-glimpses selection.

Human-captured NFOV videos ("HumanCam")

How close?



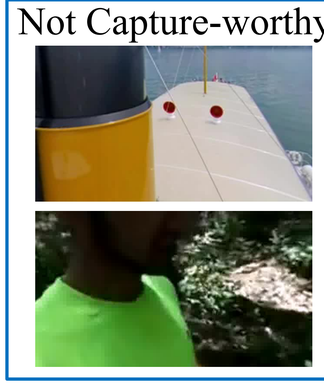
Capture-worthiness score

- Does the ST-glimpse looks human-captured?
- Implement by discriminative classifier

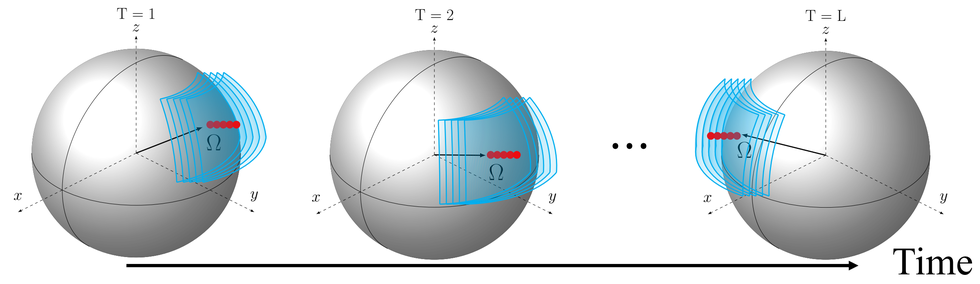
Example ST-glimpses

Scored high in capture-worthiness

Not Capture-worthy



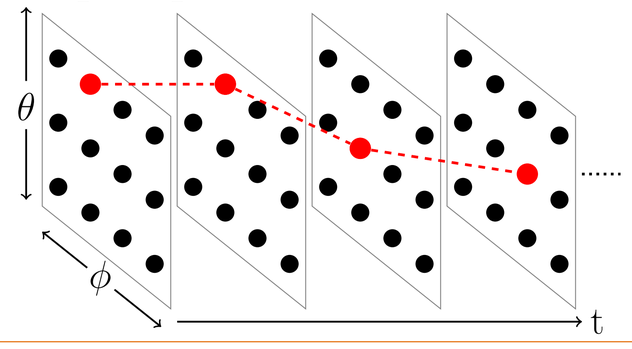
Construct Virtual Camera Trajectory



- Maximize accumulated score
- Smooth camera motion

$$|\Delta\Omega_\theta| = |\theta_t - \theta_{t-1}| \leq \epsilon_\theta,\ |\Delta\Omega_\phi| = |\phi_t - \phi_{t-1}| \leq \epsilon_\phi$$

Reduce to shortest path problem



4. Evaluation Metrics

HumanCam-based Metrics

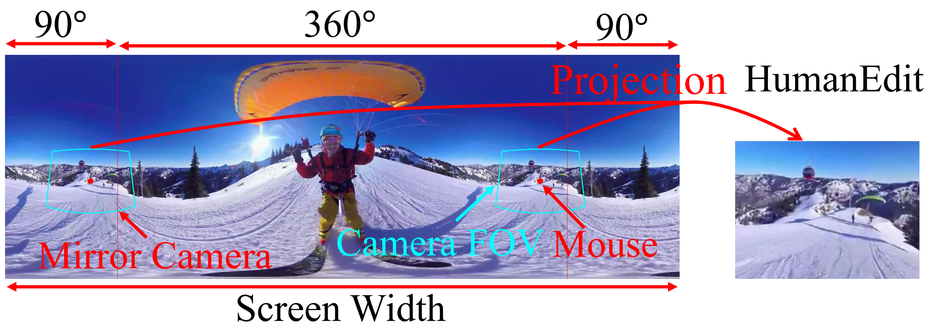
Output videos should look like human-captured videos.

- **Distinguishability**
- **HumanCam-likeness**
- **Transferability**

HumanEdit-based Metrics

Virtual camera trajectories should be similar to human-selected ones.

- **Cosine similarity**
- **FOV overlap**



5. Experiments

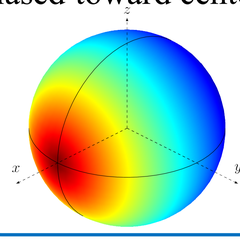
Dataset: videos cralwed from YouTube using keywords

"Hiking", "Mountain climbing", "Parade", "Soccer"

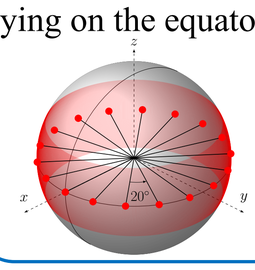
	# videos	Total length
360° videos*	86	7.3 hours
HumanCam	9,171	343 hours
HumanEdit	20#	202 minutes

Baselines

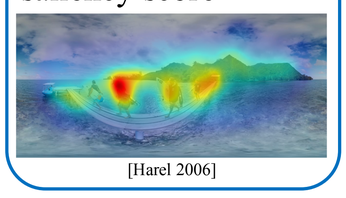
Center prior
Random trajectories biased toward center



Eye-level prior
Static trajectories lying on the equator

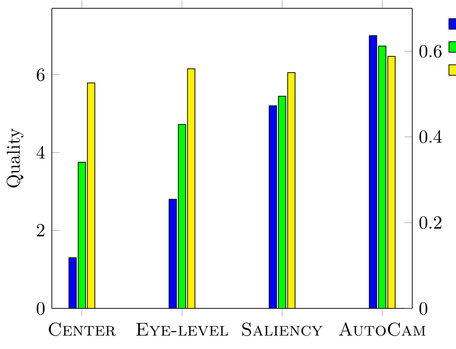


Saliency
Replace capture-worthiness with saliency score



[Harel 2006]

HumanCam-based Metrics



HumanEdit-based Metrics