Where to Look:
Navigating and Compressing 360° Video

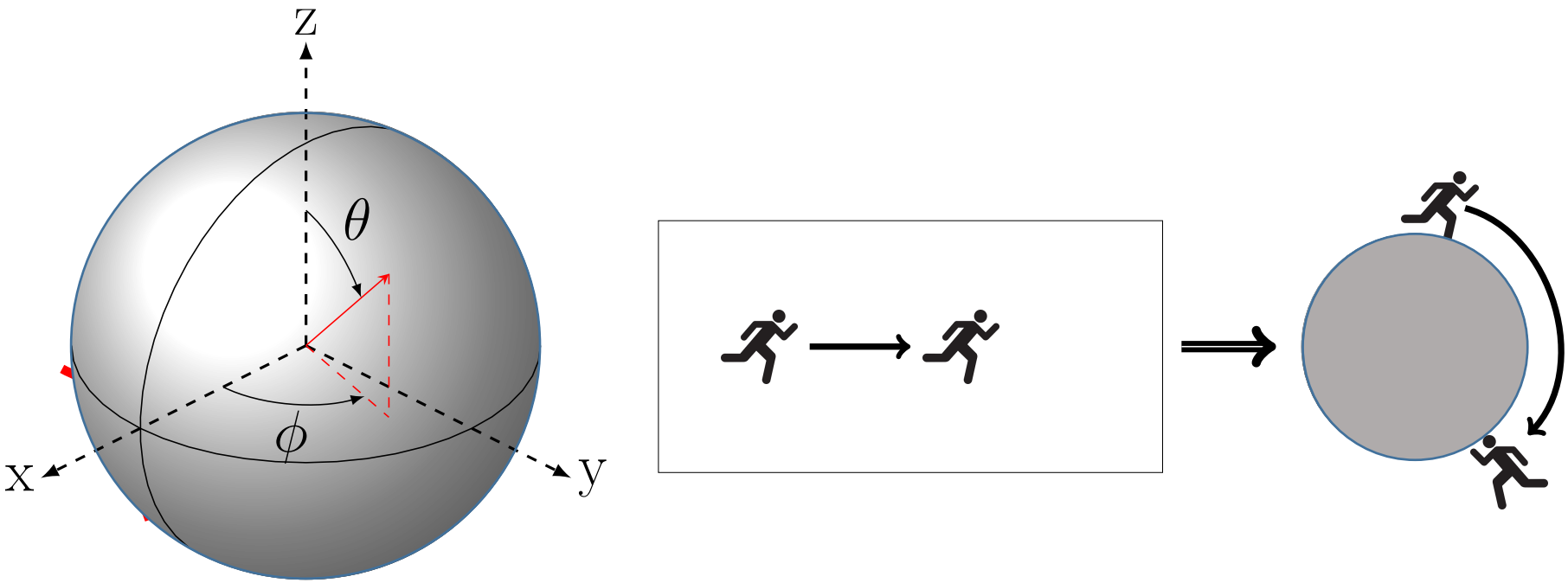
Kristen Grauman
Facebook AI Research
University of Texas at Austin

facebook artificial intelligence research

THE UNIVERSITY OF TEXAS AT AUSTIN

360° panoramic imagery

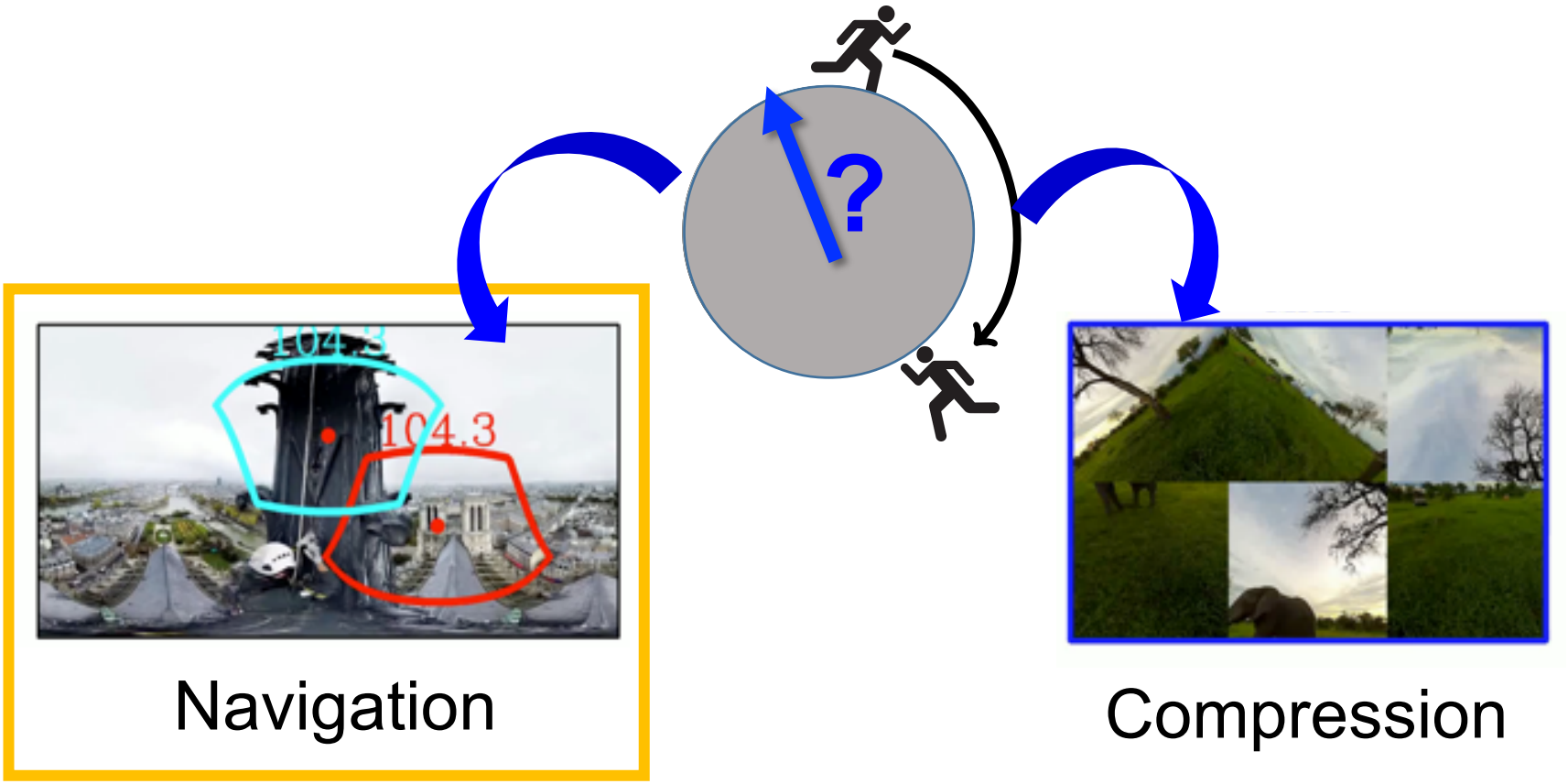
360° panoramic imagery



Field of view in spherical images bring new challenges for display and video processing.

This talk

Where to look in 360 images & video

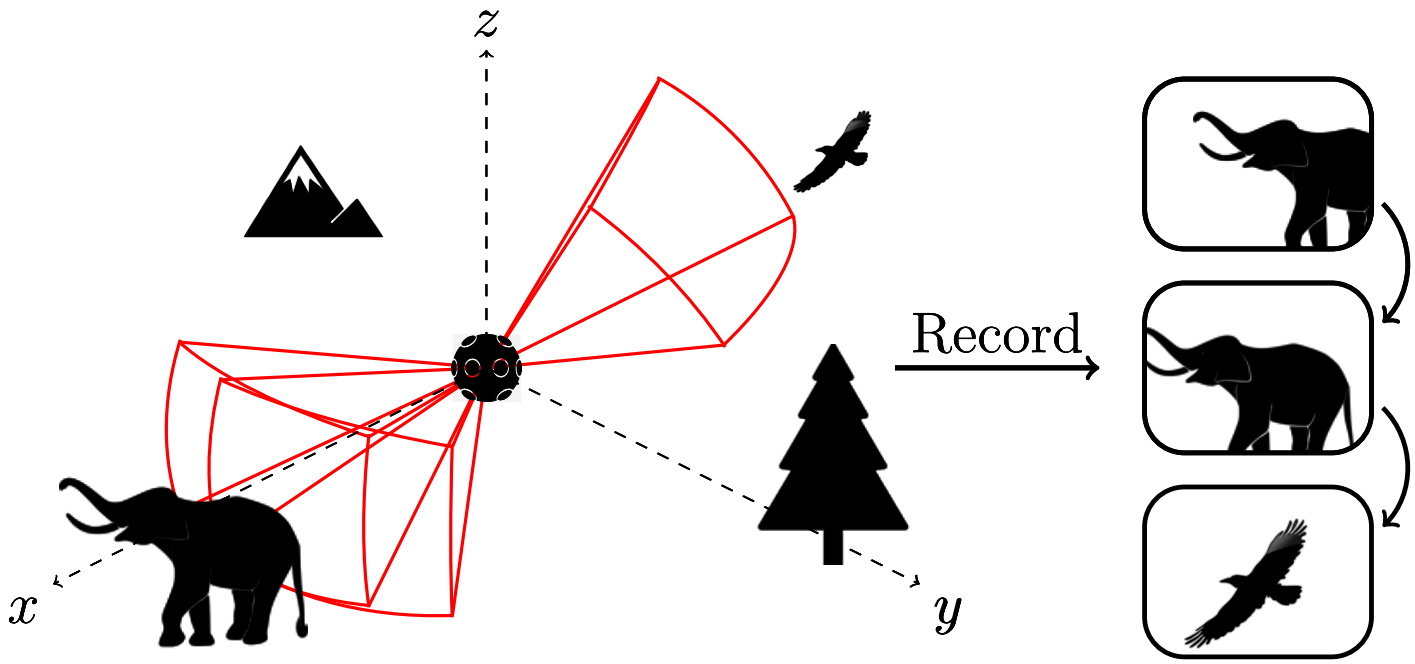


Challenge of viewing 360° videos

Control by mouse

How to find the right direction to watch?

Pano2Vid automatic videography



Pano2Vid Definition

Input: 360° video

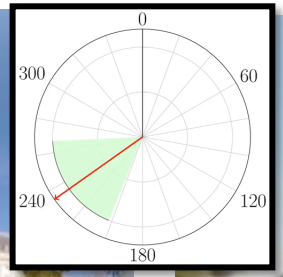
Output: natural-looking normal-field-of-view video

Task: control the virtual camera direction

[Su et al. ACCV 2016, CVPR 2017]

Pano2Vid automatic videography

Virtual camera direction



Input:
360° Video

Output:
normal-field-of-view
(NFOV) Video

[Su et al. ACCV 2016, CVPR 2017]

Our approach – AutoCam

Learn videography tendencies from unlabeled Web videos

- Diverse capture-worthy content
- Proper composition

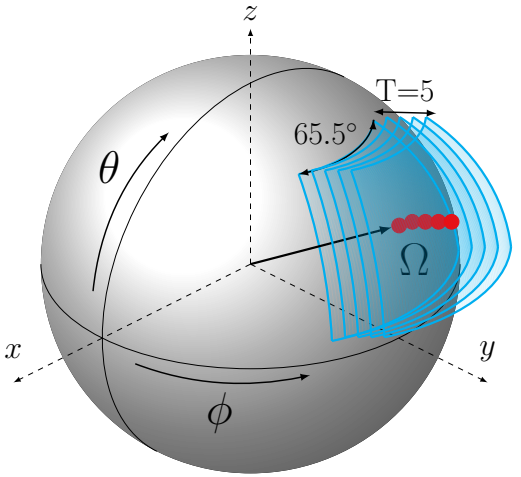
Human-captured NFOV videos ("HumanCam")

Unlabeled video

How close?

C3D features

ST-glimpses



[Su et al. ACCV 2016, CVPR 2017]

Example spatio-temporal glimpses

High capture-worthiness



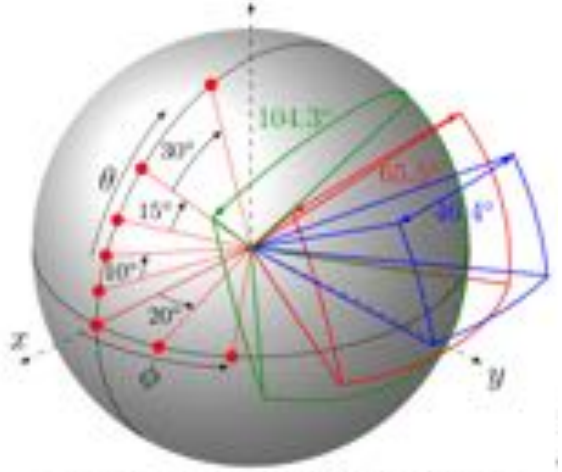



Low capture-worthiness

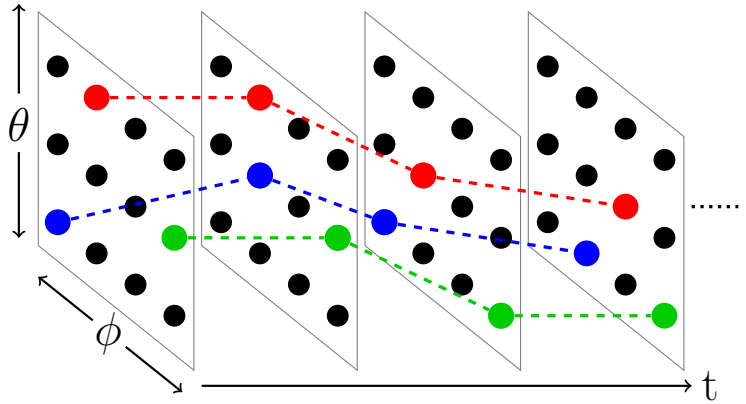


First frame of glimpses scored high/low by our approach

Construct virtual camera trajectory

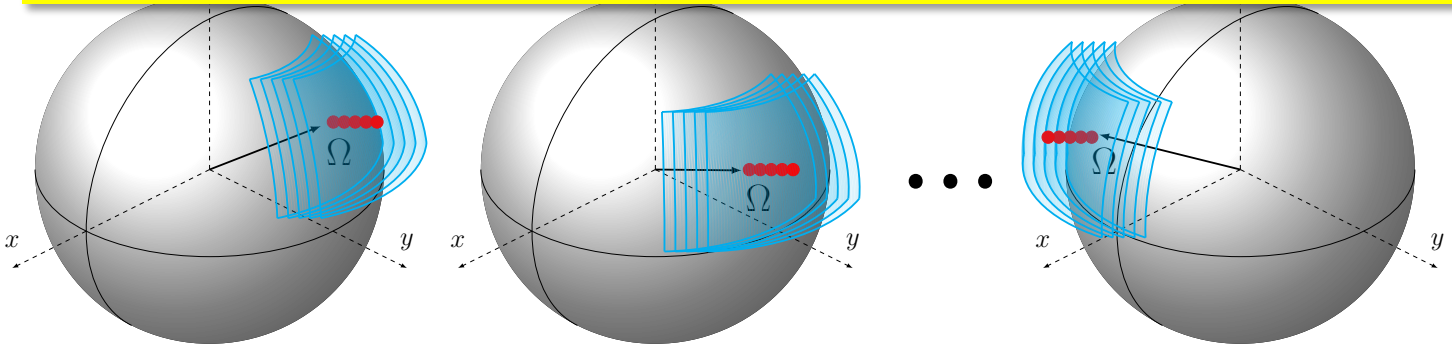


Densely sample and score glimpses



Pose selection as shortest path(s) problem

Optimize for *multiple diverse* hypotheses



Output smooth view path maximizing capture-worthiness

Datasets

- All videos crawled from YouTube using keywords:

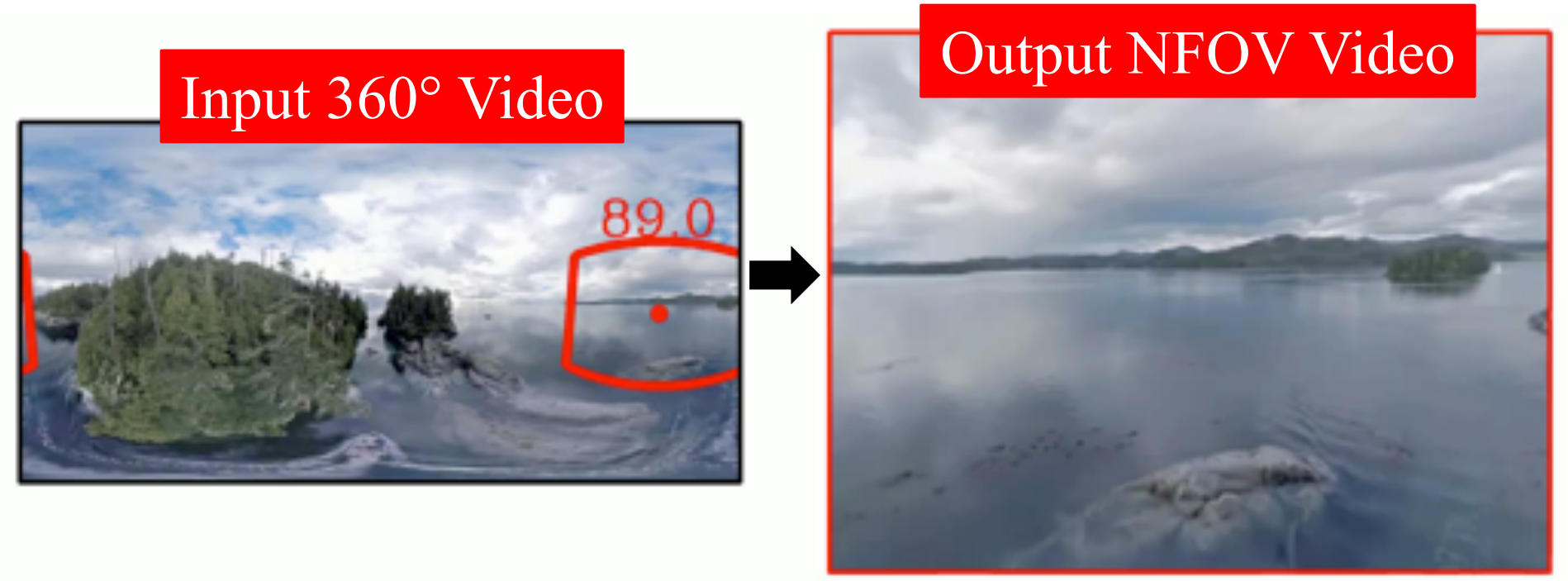
  "Hiking", "Mountain climbing", "Parade", "Soccer"

	# videos	**Total length**
360° videos	86	7.3 hours
HumanCam	9,171	343 hours

- Human-selected trajectories for evaluation only

# Editors	# Videos	# Trajectories	Total length
6	20	480	12 hours

AutoCam results



Automatically select FOV and viewing direction

[Su & Grauman, CVPR 2017]

AutoCam results

Automatically select FOV and viewing direction

[Su & Grauman, CVPR 2017]

AutoCam results:
Multiple diverse hypotheses

Input Video & Cam. Trajectory

Output Videos

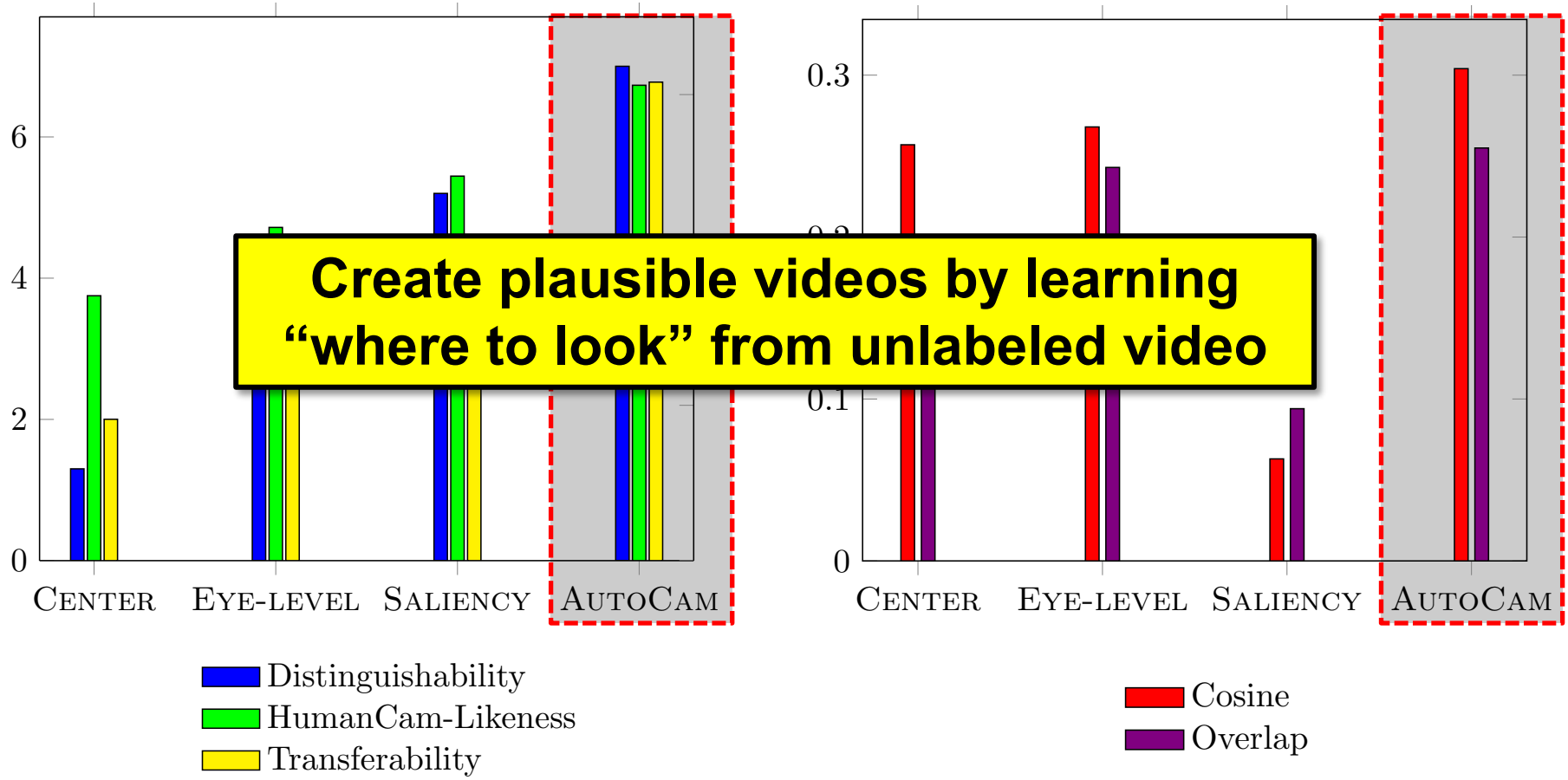
Hypothesis 1

Hypothesis 2

AutoCam results

Similarity to user-uploaded standard web videos

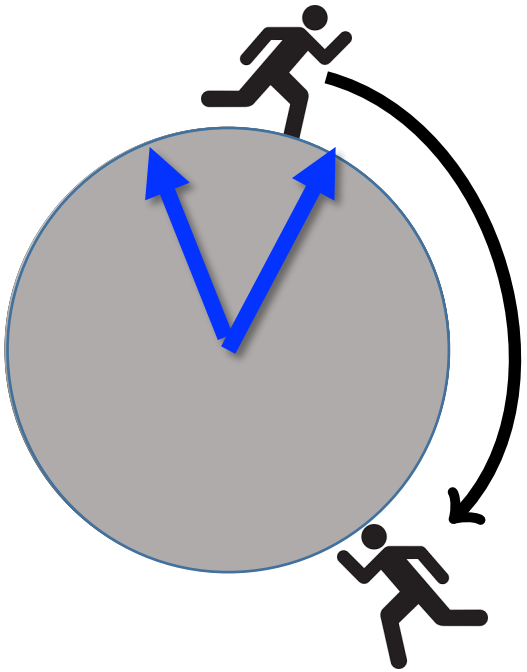
Similarity to human-selected camera trajectories



Create plausible videos by learning "where to look" from unlabeled video

[Su et al. ACCV 2016, CVPR 2017]

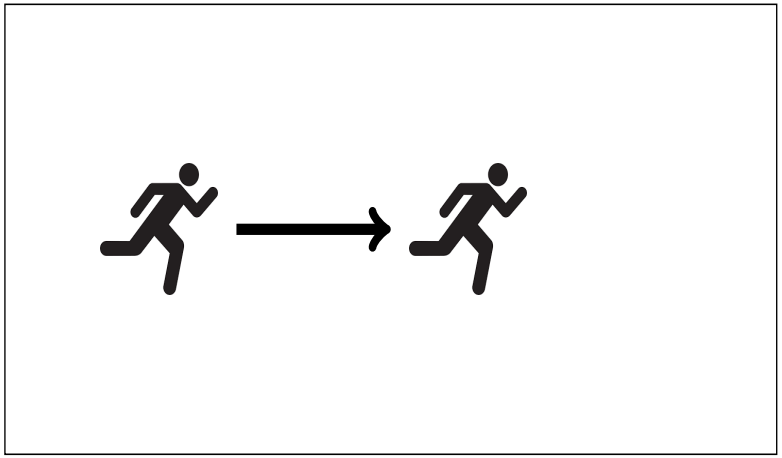
The “where to look” task for display



Carving a “normal” field of view

+ focus, undistorted view

- always miss something

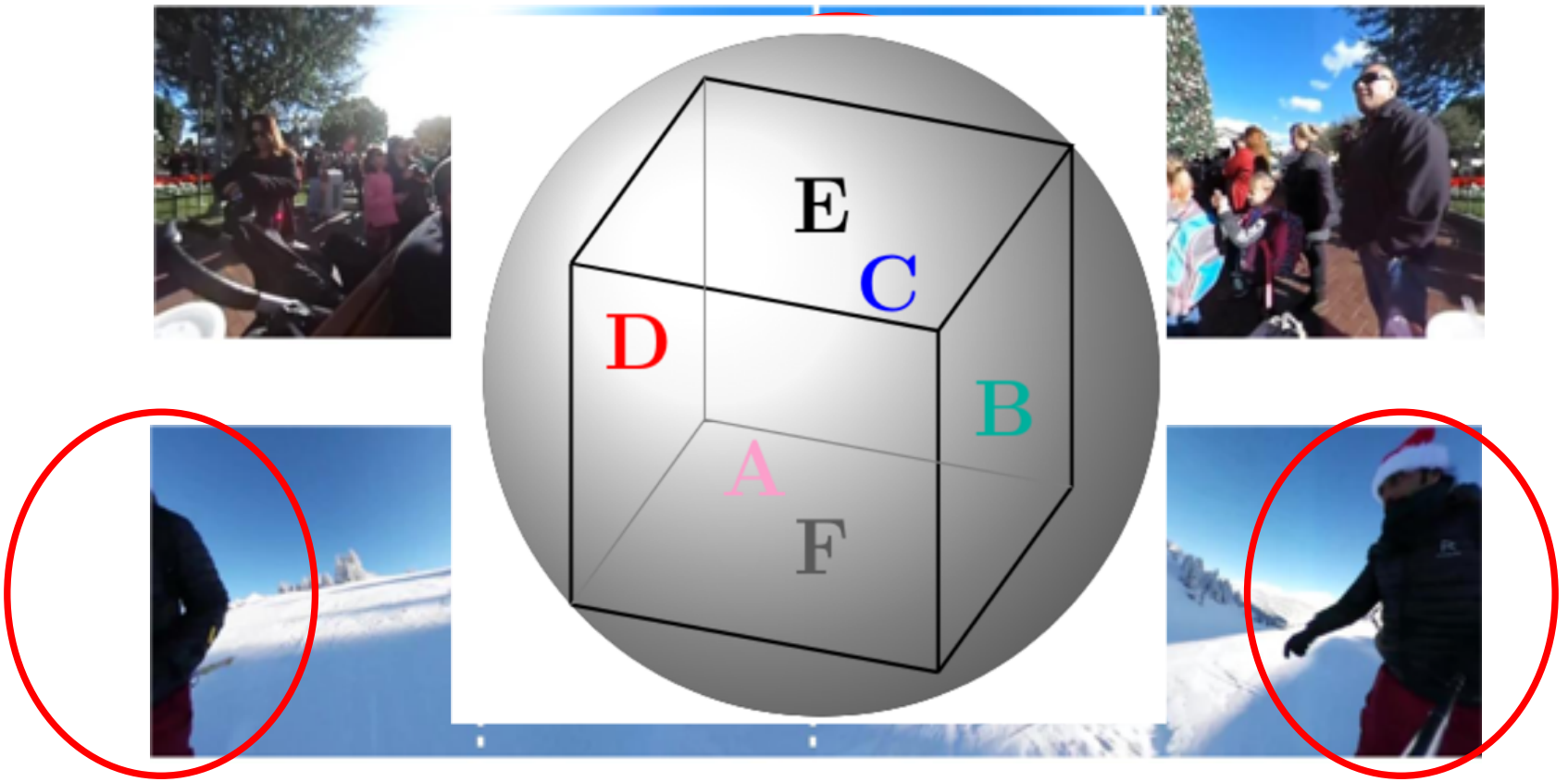


Selecting a good projection

+ see almost everything

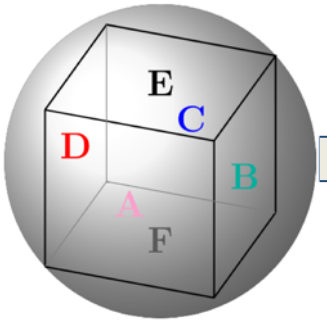
- always warp something

Cubemaps



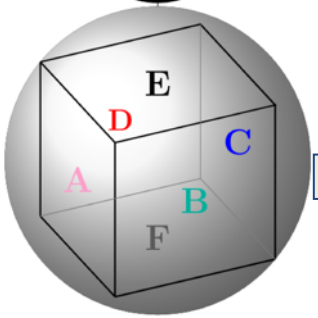
Problem: Object integrity damaged when same object is projected onto different cube faces.

Our idea: Snap angles

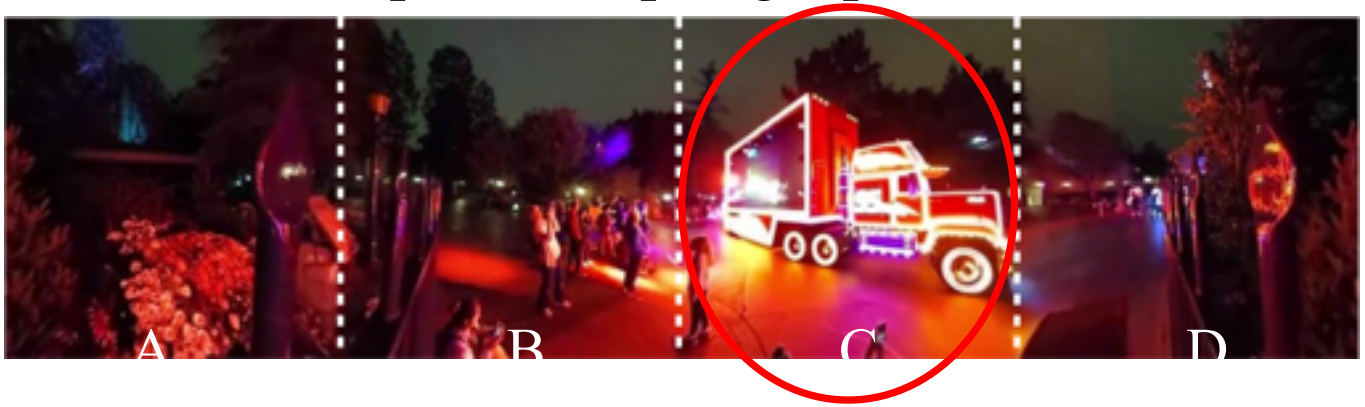


Default 360º cubemap

Rotate!



Proposed snap angle prediction



How to predict the viewing angle?

Xiong & Grauman, ECCV 2018

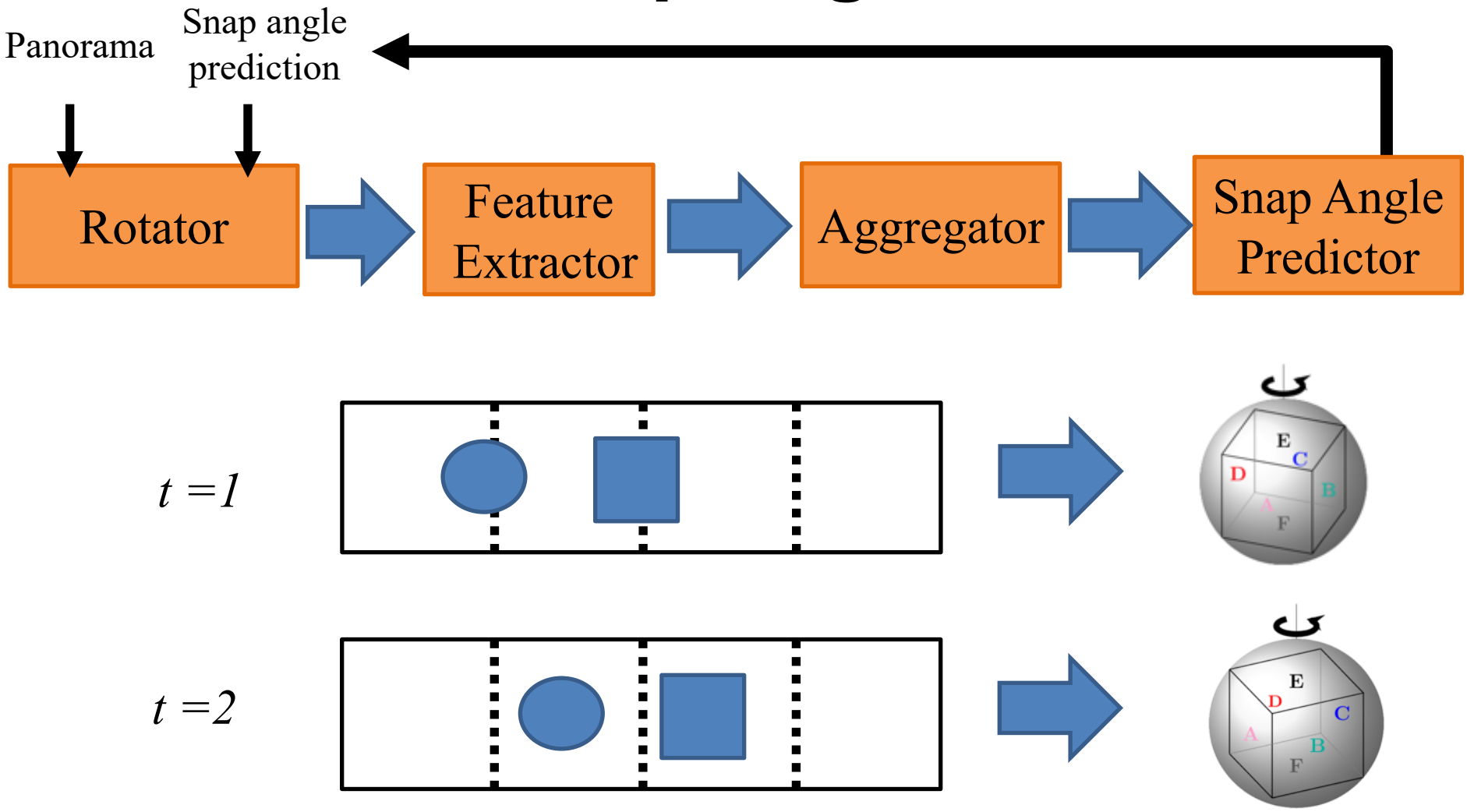
Snap angles formulation

Objective: minimize area of foreground objects near or on cube boundaries.

Pixel objectness [Jain et al. PAMI 2018]

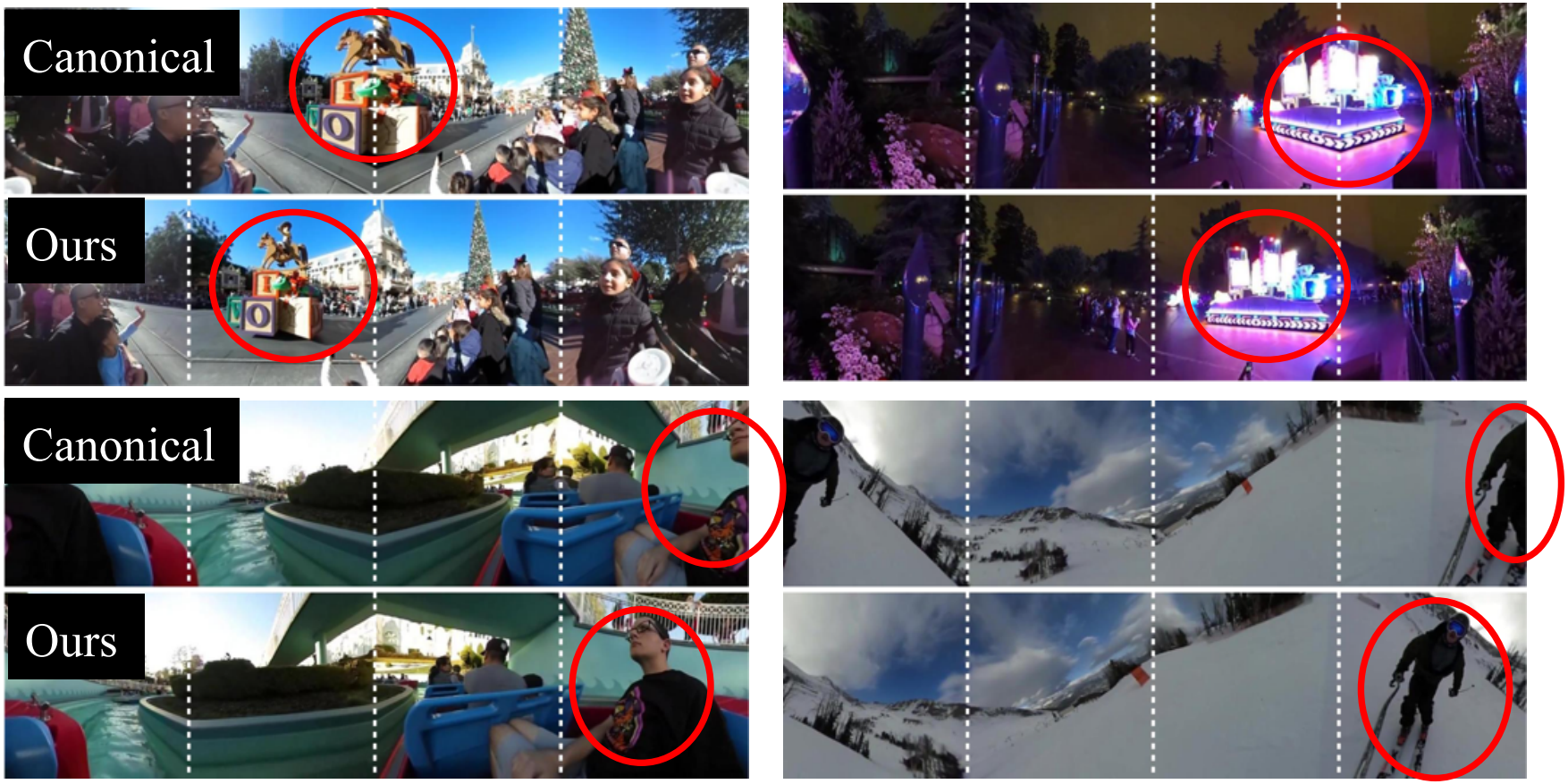
Approach: time-budgeted sequential decision process to rapidly infer best snap angle

Snap angles



Recurrent neural network learns sequence of rotations that minimize foreground disruption

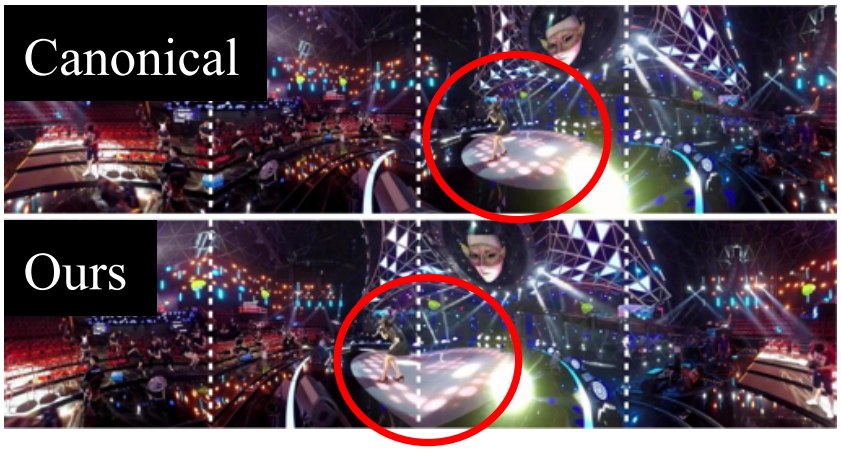
Snap angle results

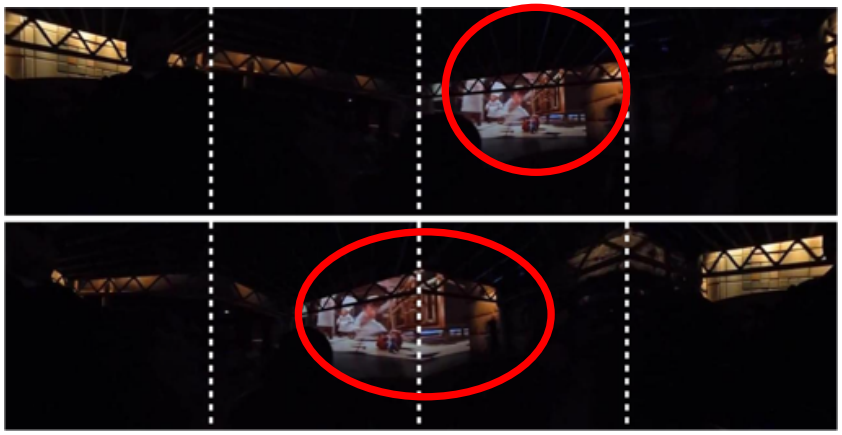


Foreground objects preserved in same cube face

Snap angle results

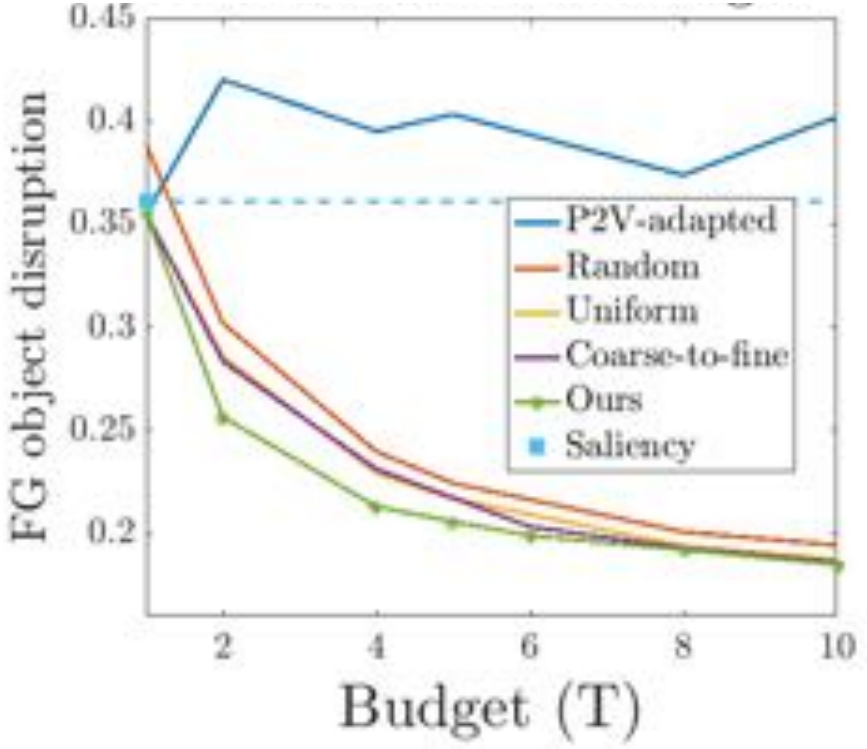
Failure cases



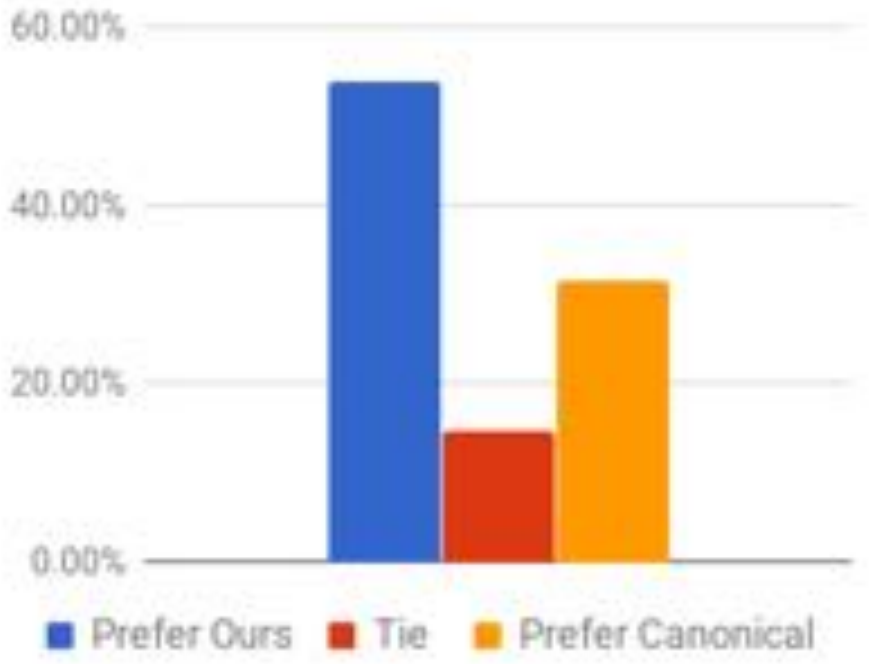


Affected by foreground errors

Snap angle results



Efficient angle selection

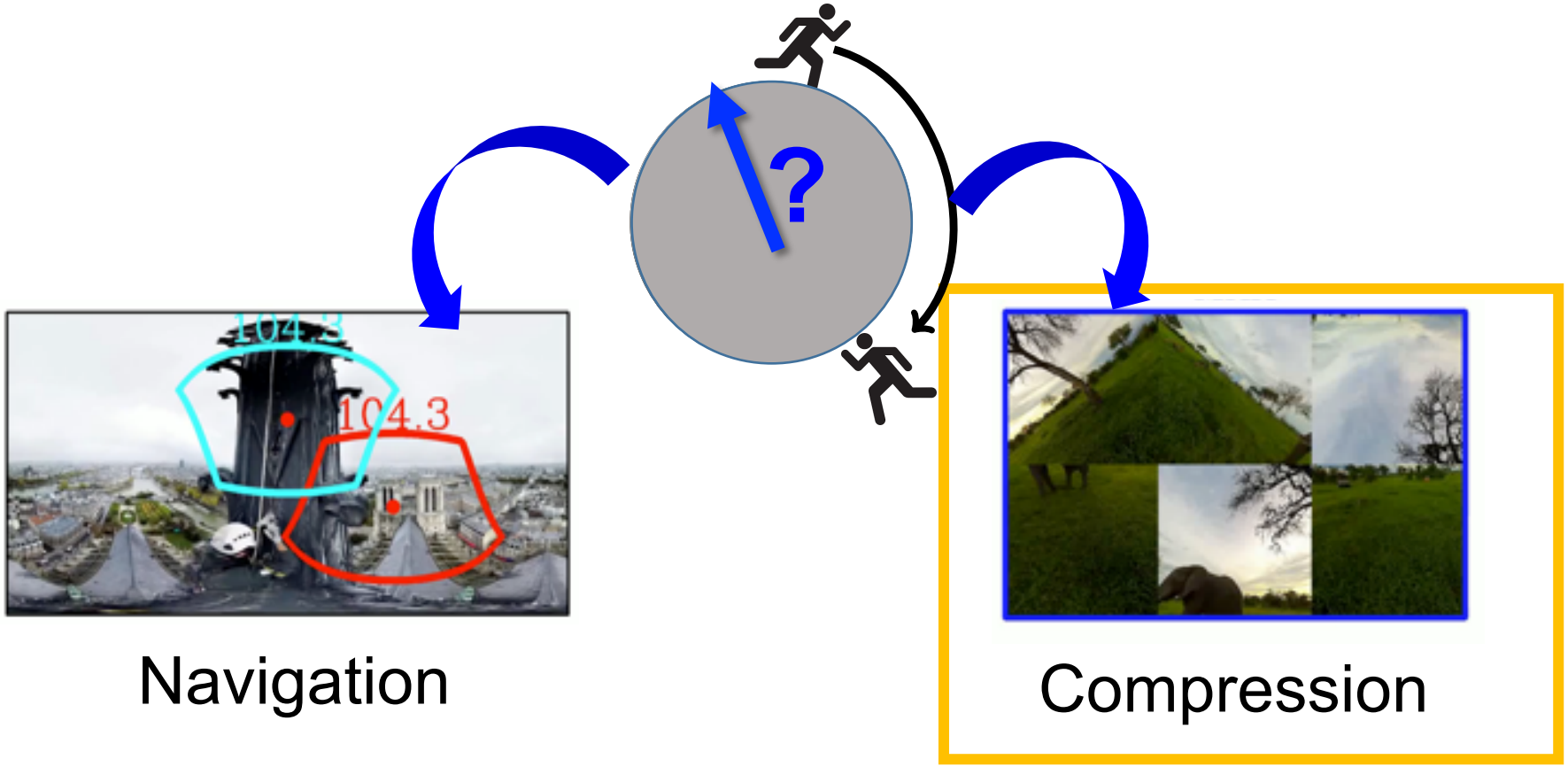


Human-perceived quality

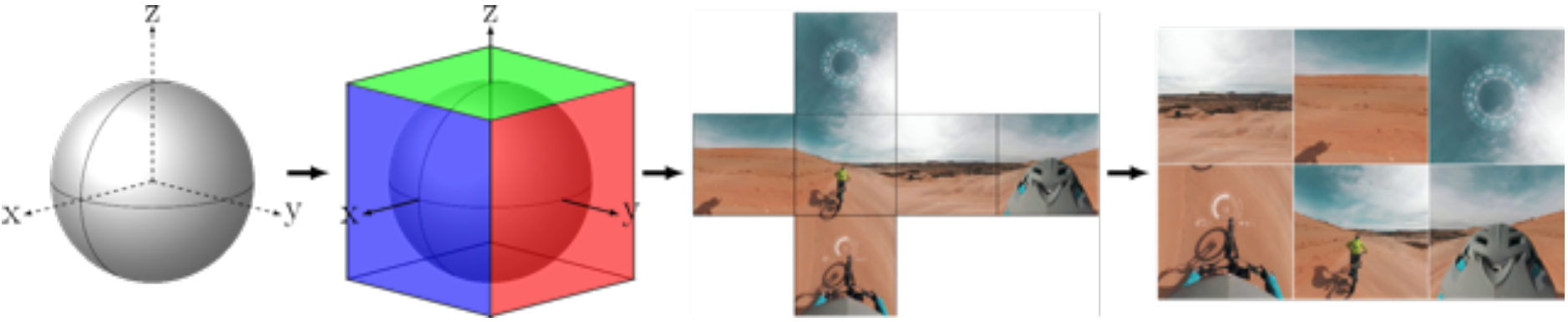
Xiong & Grauman, ECCV 2018

This talk

Where to look in 360 images & video



How to compress a 360 video?

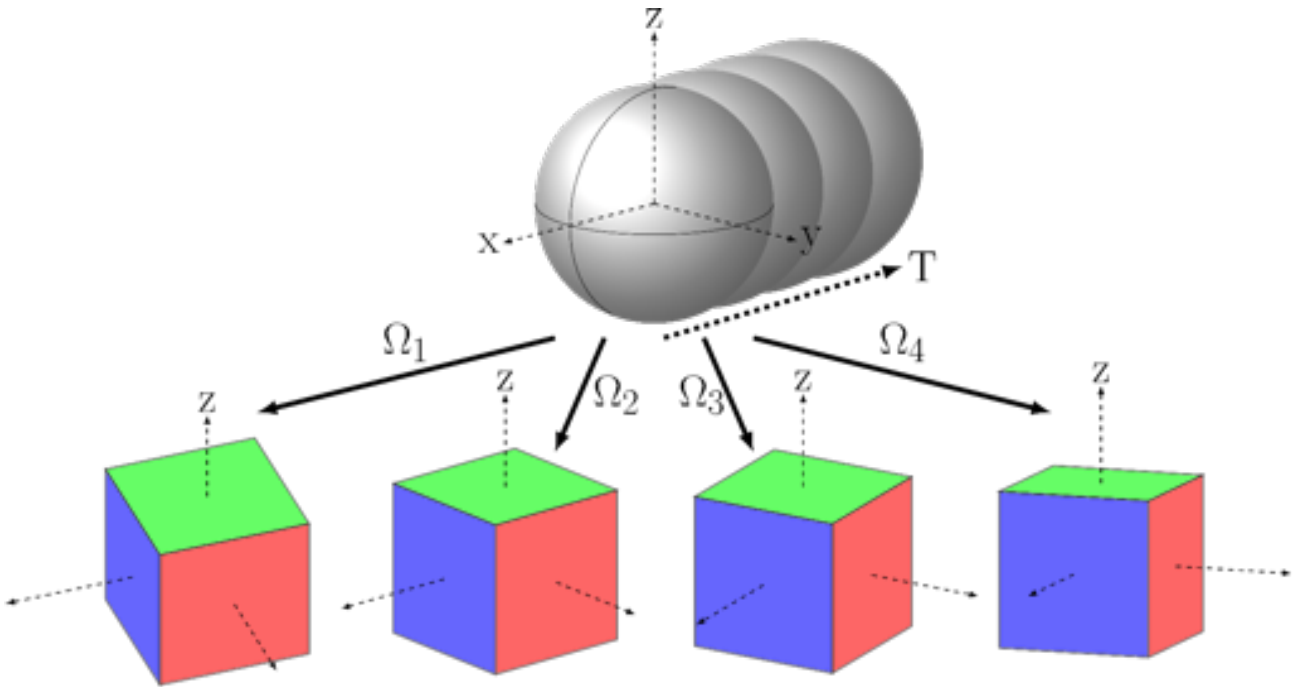


Cubemap projection

Restacked

Status quo – apply usual encoders to cubemap

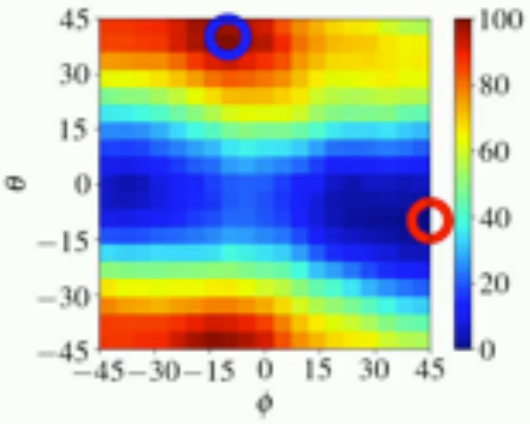
Problem: 360 video isomers



- Video content is invariant to projection axis
- However, the encoded bit-streams are not

[Su & Grauman, CVPR 2018]

Problem: 360 video isomers



Video size vs.
cube rotation angle

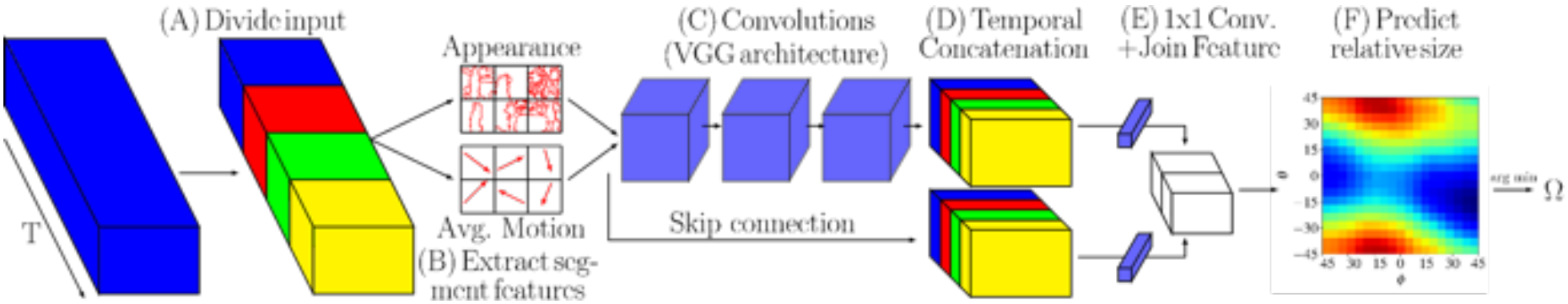
MIN

MAX

- Video content is invariant to projection axis
- However, the encoded bit-streams are not

[Su & Grauman, CVPR 2018]

Our idea: Compressible 360 isomers



Given video, predict most compressible isomer (angle)

[Su & Grauman, CVPR 2018]

Compressible 360 isomer results

Predicted angle vs. file size heatmap

[Su & Grauman, CVPR 2018]

Compressible 360 isomer results

Max Video Size
0% Reduction

Min Video Size
100% Reduction

	H264	HEVC	VP9
RANDOM	50.75	51.62	51.20
CENTER	74.35	63.34	72.92
OURS	82.10	79.10	81.55

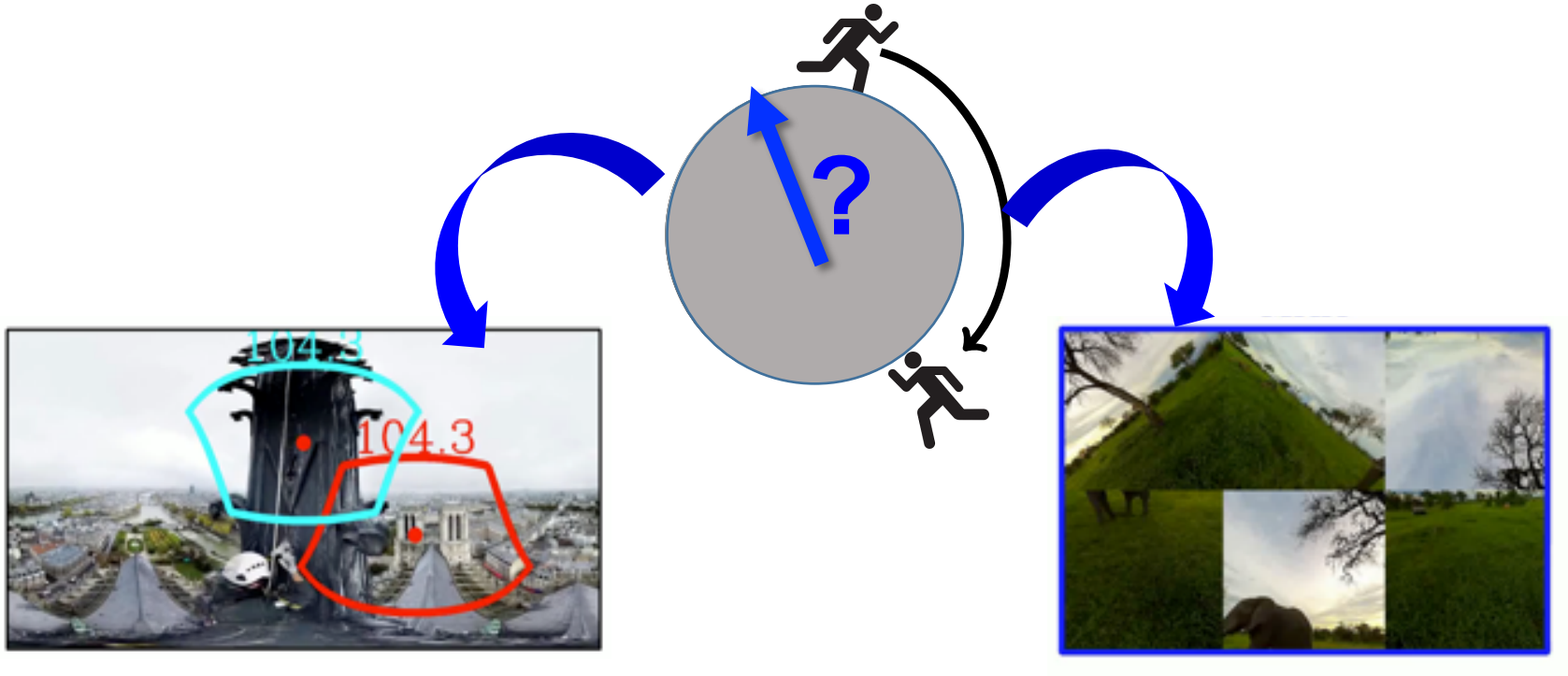
% size reduction achieved

Uses < 0.3% the computation of exhaustive search

[Su & Grauman, CVPR 2018]

This talk

Where to look in 360 images & video



Navigation

Compression

Kernel Transformer Networks

How to translate a CNN to spherical images?

[Su & Grauman, CVPR 2019]

Summary

Kristen Grauman
FAIR & UT Austin

- New challenges for 360 display and video processing
- Our idea - Intelligent selection of view/orientation
  - For better human consumption (AutoCam, Snap Angles)
  - For better video compression

Yu-Chuan Su

Bo Xiong

AutoCam
[ACCV 2016, CVPR 2017]
Snap Angles
[ECCV 2018]
Compressible isomers
[CVPR 2018]