

Detecting Engagement in Egocentric Video

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

<http://vision.cs.utexas.edu/projects/ego-engagement>



1. Engagement in Egocentric Video

Motivation: people do not always engage with what they see and pay different levels of attention to the environment

Goal: given an egocentric video, we want to predict **when** the camera wearer is engaged with what he sees.



Definition of Engagement

The recorder is attracted by some object(s), and he interrupts his ongoing flow of activity to purposefully gather more information about the object(s).

Applications Behavior analysis Camera control VR display



Engagement is different from saliency: Previous work [Harel '06, Itti '09, Rudoy '13, ...] on visual attention focuses on **where** the people look but ignores **when** people are engaged.

2. UT Egocentric Engagement (UT EE) Dataset

We collect videos in three **browsing** scenarios:



Video Statistics

- 27 videos
- 9 recorders
- 14 hours total length

Engagement Annotation

Frame-level annotation with MTurk. Each video is labeled by **10 Turkers**. Ground truth is determined by majority vote.

3. Data Analysis

We collect **3 hours** of recorder self-annotation to verify the third person annotation.

			Frame F_1	Interval F_1	
				Boundary	Presence
Turker	vs. Consensus		0.818	0.837	0.914
	vs. Recorder		0.589	0.626	0.813
Random	vs. Consensus		0.426	0.339	0.481
	vs. Recorder		0.399	0.344	0.478

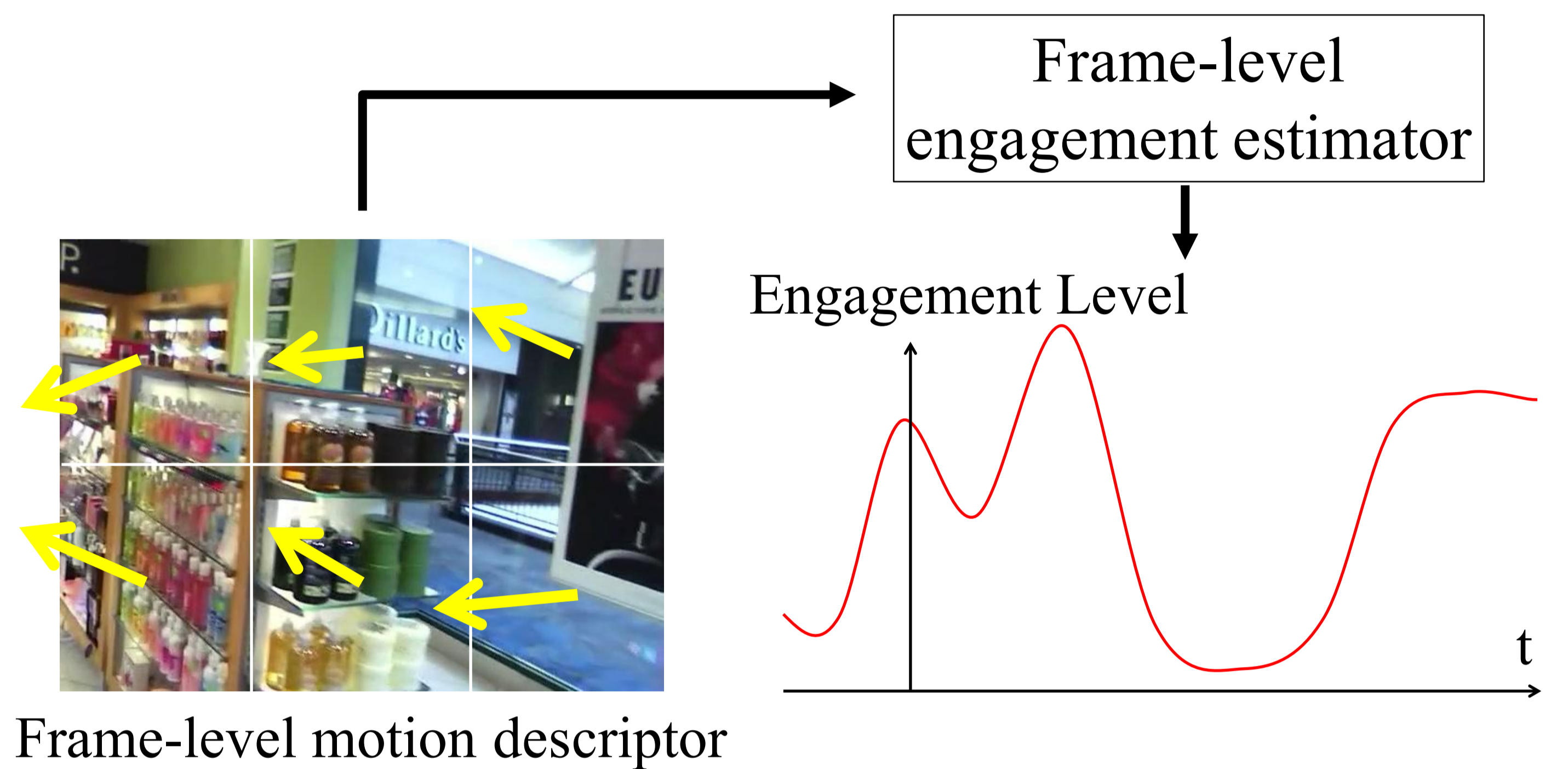
Engagement is predictable from egocentric video!

4. Predict Engagement from Motion

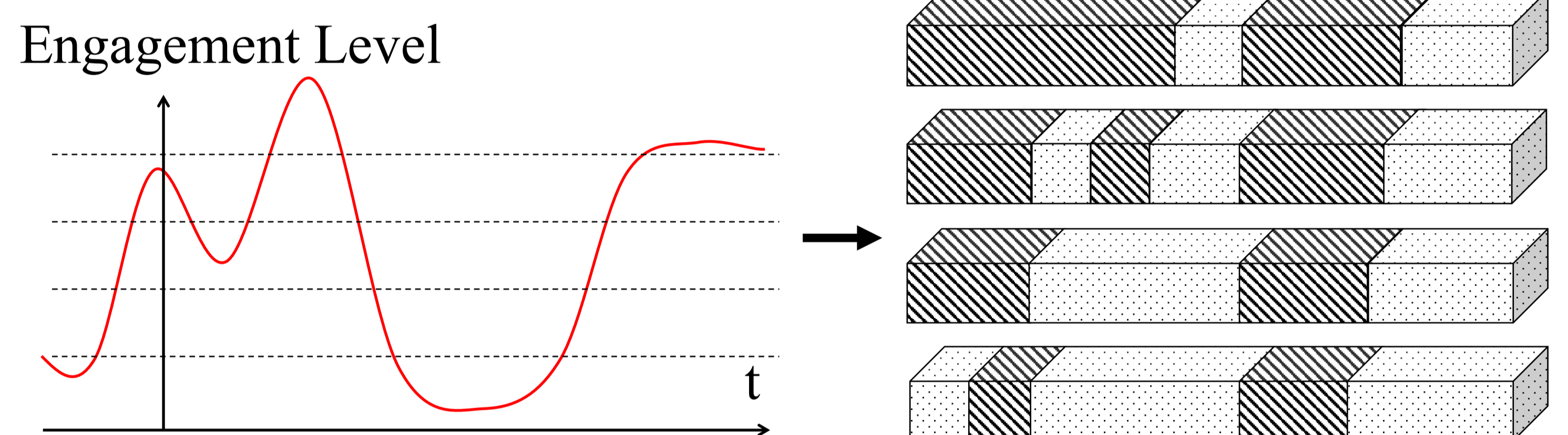
Challenge of engagement detection

- Diverse visual content
- Being engaged \neq being static
- Duration of engagement varies significantly

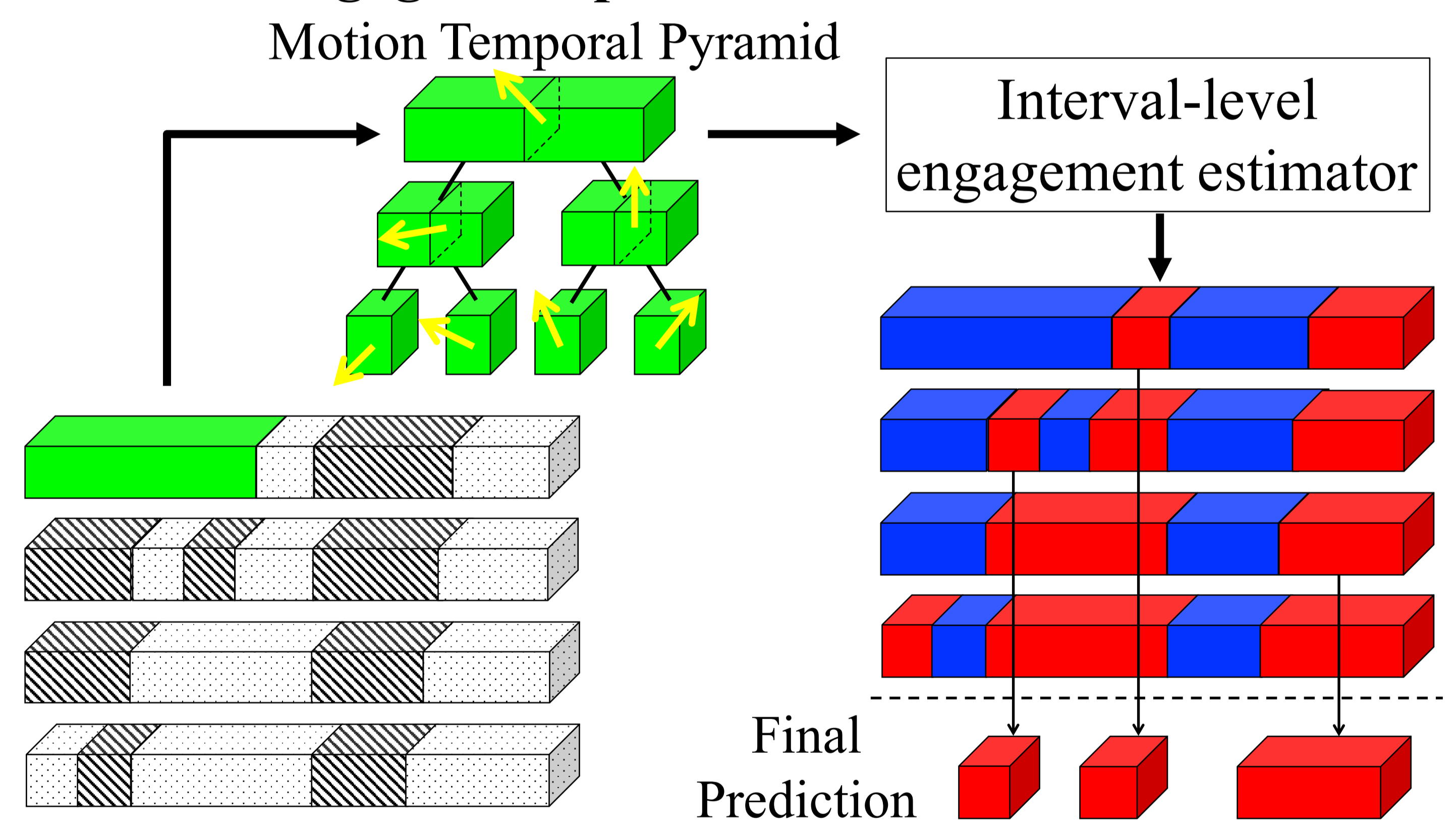
1. Estimate frame-wise engagement



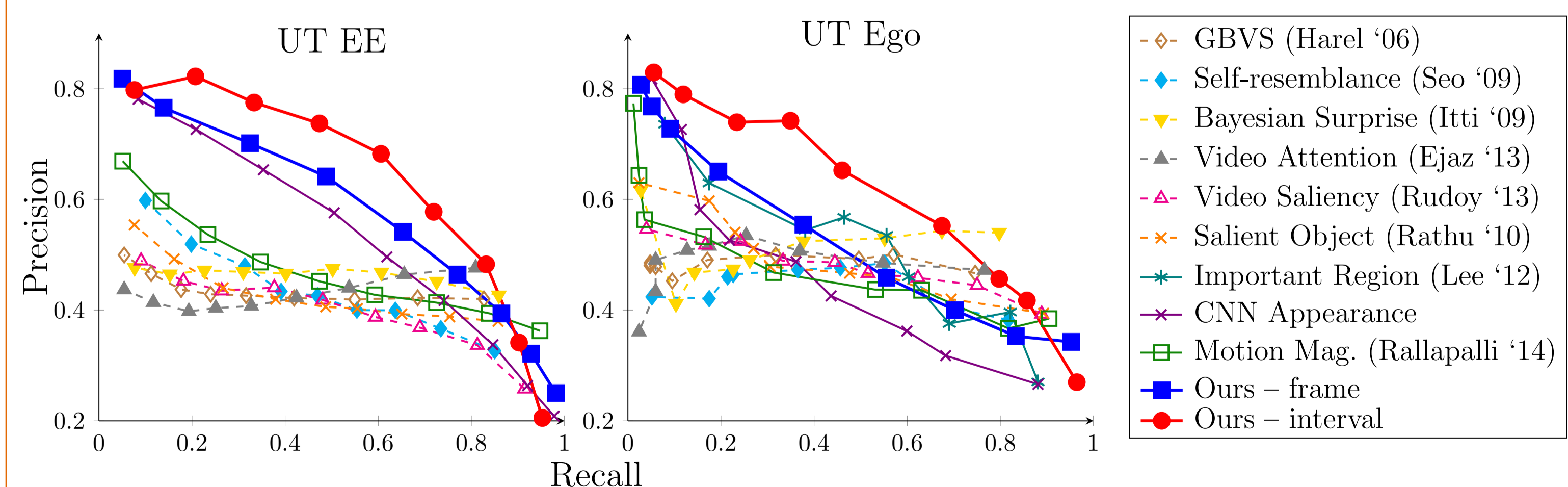
2. Generate interval hypotheses



3. Estimate engagement per interval



5. Experiments



- Our method performs the best in all settings
- Interval hypothesis has clear positive impact
- Appearance feature does not generalize well (UT Ego)
- Saliency/Motion Mag. performs poorly
- We outperform Important Region without train on UT Ego