Dynamic Object Scanning: Object-Based Elastic Timeline for Quickly Browsing First-Person Videos

CHI 2018 Engage with CHI

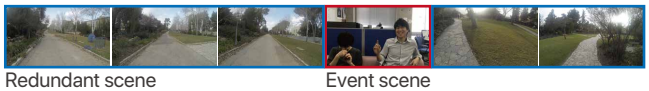
Seita Kayukawa[1] Keita Higuchi[2] Ryo Yonetani[2] Masanori Nakamura[1] Yoichi Sato[2] Shigeo Morishima[3]

[1]Waseda University [2]University of Tokyo [3]Waseda Research Institute for Science and Engineering

Introduction

First-person videos captured by wearable cameras

- Large and diverse collection of **long and untrimmed** videos
- Important events can be distributed sparsely



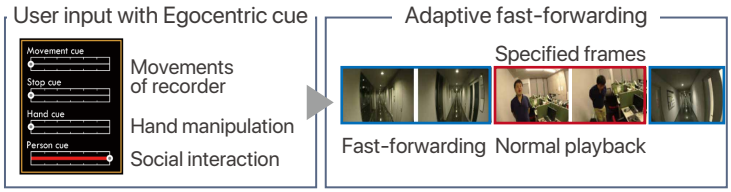
Redundant scene Event scene

Our Goal

To present adaptive video fast-forwarding technique that helps users to browse long and untrimmed videos quickly

Related Work

Adaptive Video Fast-forwarding based on User Input[1]

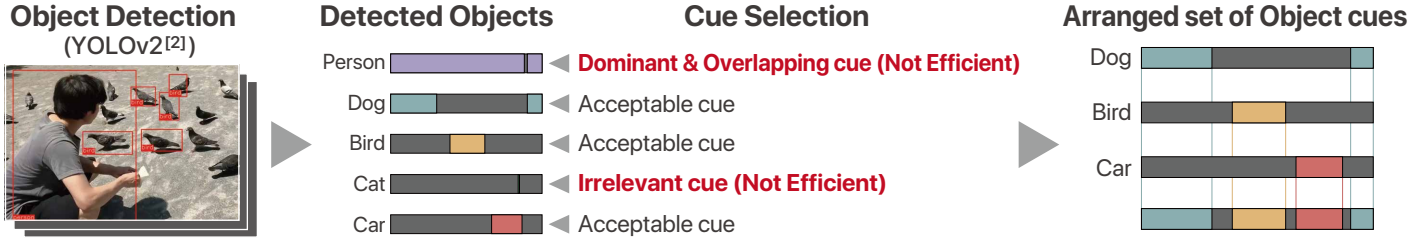


Limitation: The type of cues have been **fixed for any given video**

▶ Limits the variety of videos that can get the benefit of Egocentric cue

Our Approach

Key Idea: To generate a efficient set of cues arranged adaptively tailored to the contents of a given video



Greedy Algorithm

Objective function

$$F(\mathcal{C}) = \boxed{A(\mathcal{C})} - \boxed{B(\mathcal{C})}$$

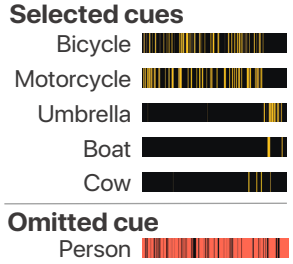
$\mathcal{C} = \{c_1, \ldots, c_N\}$: set of detected objects

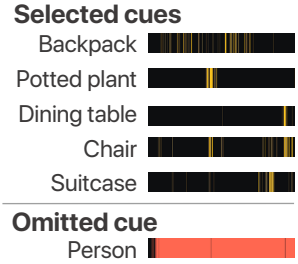
$A(\mathcal{C})$: the number of frames where at least one of the categories in $\mathcal{C}$ is observed
To avoid an **irrelevant category** and a **temporally-overlapping category**

$B(\mathcal{C})$: the number of frames with the object category observed most frequently
To avoid a **temporally-dominant category**

Cue Selection Results

Applied our cue selection algorithm to diverse scenes





Pedestrians were **detected in nearly every frames** ▶ Person was **not selected** as cue

Our algorithm selects a set of cues while excluding temporally dominant categories

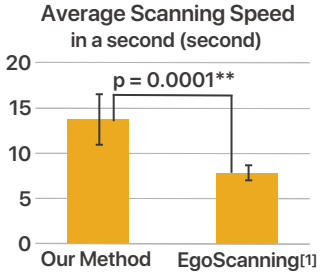
User Study

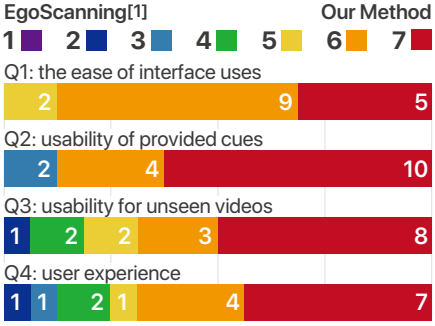
Task: Finding predefined event from long first-person videos

Participants: 16 university students

Our Interface: Object cues selected by our method

Baseline (EgoScanning[1]): **Fixed** Egocentric cues





Feedback

- Object cues helped several users to **infer the contents of given videos**.
- Object cues are useful to find important events since they often **emphasized only a limited part of videos**.

References [1] K.Higuchi et al., "EgoScanning: Quickly Scanning First-Person Videos with Egocentric Elastic Timeline", CHI 2017 [2] J.Redmon, "YOLO9000: Better, Faster Stronger", CVPR 2017