Learning to Predict Gaze in Egocentric Video

Yin Li, Alireza Fathi, James M. Rehg
School of Interactive Computing, Georgia Tech

Motivation

“Nothing is more powerful than an idea whose time has come” – Victor Hugo

- Gaze is a very important signal
- Predicting the camera wearer’s gaze using egocentric cues
- Understanding first person’s behavior using predicted gaze

Method

- Modeling the first person’s head-eye / hand-eye coordination
- Only use egocentric cues, e.g. hand pose, head movement
- A temporal dynamic model for fixations

Head-Eye Coordination

- Large head motion is always accompanied by a large gaze shift
- Linear correlation of head motion and gaze shift in horizontal direction

Hand-Eye Coordination

- Peak of gaze distributions around hands, where manipulations are most likely to happen

Object

- Object involved in current task
- 94.3% objects with fixations
- Improved object segmentation (+3%)

Action

- Action recognition of 25 classes using predicted gazes > 32.8%
- Action recognition using human gaze s > 49%

Acknowledgement:

Potion of this work was funded in part by the National Science Foundation under Grant No.CCF-1029679. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation (NSF).

Contact:

Yin Li (yli440@gatech.edu) Webpage: yinli.cvpr.net

Reference:

Learning to Predict Gaze in Egocentric Video, ICCV 2013