GazeCapture

Links

- Website: http://gazecapture.csail.mit.edu/
- Paper: https://people.csail.mit.edu/khosla/papers/cvpr2016_Khosla.pdf
- GitHub: https://github.com/CSAILVision/GazeCapture/tree/master/pytorch
- Using on infant: http://vision01.csail.mit.edu/gaze/scratch2/recasens/opengaze/videos_supp/baby_output.mp4

Notes

Goal

- Make eye-tracking more pervasive by building eye tracking software that works on commodity hardware such as mobile phones and tablets, without the need for additional sensors or devices

About GazeCapture

- The first large-scale dataset for eye tracking
  - ~1500 subjects from a wide variety of backgrounds, recorded under variable lighting conditions and unconstrained head motion
  - Collect data by crowdsourcing - more variation, less costly, more efficient (than inviting participants to lab)
    - Scalability: hybrid approach of combining the scalable workforce of crowdsourcing platforms together with the design freedom provided by building custom mobile apps -- created iOS application called GazeCapture and used AMT as a platform for recruiting people to use app
    - Reliability (making sure workers are doing what they are supposed to do)
    - Variability
      - Crowdsourcing leads to large variability in pose, appearance, and illumination
      - Tell workers to continuously move head and relative distance to phone
      - Force workers to change orientation of mobile device
  - Data uploaded as individual frames rather than a video (to avoid compression artifacts)
- Mobile-based eye tracking dataset

iTracker

- CNN for eye tracking
- Trained using GazeCapture
- Achieves a significant reduction in error over previous approaches while running in real time on a modern mobile device
- Does not rely on any pre-existing systems for head pose estimation or other manually-engineered features for prediction
- Trained with crops of both eyes and the face
- Achieves state-of-the-art performance in terms of accuracy
- Problem: size of inputs and number of parameters make it difficult to use in real-time on a mobile device
    - Solution: apply ideas from the work on dark knowledge to train a smaller and faster network that achieves real-time performance on mobile devices with a minimal loss of accuracy
        - Applied dark knowledge to reduce model complexity, and thus, computation time and memory footprint
- Input: face grid (image of face together with its location in the image) and image of the eyes
- Using the model can: infer the head pose relative to the head, and infer the pose of the eyes relative to the head
    - Model uses this info to infer the location of gaze
- Designed a unified prediction space to train a single model using all the data
    - Leveraged the fact that the front-facing camera is typically on the same plane as, and angled perpendicular to, the screen
    - Afterwards, fine-tuned the network to each device and orientation
        - Useful in dealing with the unbalanced data distribution between mobile phones and tablets

Results

- Average error of ~2 cm without calibration
- Error reduced to 1.8cm through calibration
    - Performance decreases slightly when given few points for calibration (due to overfitting)

On infants

- General direction of gaze is accurate
- Box size is not steady - sometimes increases significantly in size
    - Some distances (?) unstable (1:50/2:00) - iTracker has trouble detecting the infant’s face, box on average includes a larger area (includes significant area below face)
- Looking at extreme directions (e.g. to the right) increases the size of the box

- Sometimes registers a “different” face that includes both the infant’s face and the back of the parent’s head, with a gaze in a completely different direction than the infant’s
    - Gaze arrow of other “face” may start from same location as the infant’s, or a different location within the infant’s face
- Gaze is not steady - tends to move around a lot even when infant is steadily looking in one direction

Comments

- Relies on crowdsourced data from app on iPhones - data would consist of people old enough to own a phone