Eye Tracking for VR and AR

Organizers:

- Robert Cavin, Facebook Reality Labs, robert.cavin@oculus.com
- Jixu Chen, Facebook, jixu.chen@oculus.com
- Ilke Demir, DeepScale, idemir@purdue.edu
- Stephan Garbin, Facebook Reality Labs, stephan.garbin@oculus.com
- Oleg Komogortsev, Texas State University, oleg.komogortsev@gmail.com
- Immo Schuetz, Facebook Reality Labs, ischtz@oculus.com
- Abhishek Sharma, Gobasco Pvt. Ltd., abhisharayiya@gmail.com
- Sachin S. Talathi, Facebook Reality Labs, sachin.talathi@oculus.com

1.0. Topics

Virtual (VR) and Augmented Reality (AR) has garnered mainstream attention with the advent of products such as the Oculus Rift, HTC Vive, Microsoft HoloLens, and Magic Leap One. However, these products have yet to find broad adoption by consumers. Mass market appeal for these products may require revolutions in comfort, utility and performance. These revolutions can, in part, be enabled by measuring where an individual is looking, where his/her eyes are, and the state of his/her eye, colloquially "eye-tracking."

In order to create a 3D immersive experience, all VR and AR devices spend significant energy in rendering immersive 3D environments. Tracking the position and orientation of the eye as well as gaze direction unlocks novel display and rendering architectures that can substantially alleviate the power and computational requirements to render 3D environments. Furthermore, eye-tracking enabled gaze prediction and intent inference can enable intuitive and immersive user experiences adaptive to the user requirements in his/her interaction with the virtual environment.

These applications require eye-tracking to reliably work **all the time for all individuals under all environmental conditions** within the power and computational constraints imposed by the form-factor of VR and AR devices. These generalizability requirements are critical for any tracking technology.

A key milestone toward this objective is a dataset that couples tracking-system input to ground-truth eye state and eye gaze. Since it is difficult to instrument an experiment to have explicit control over eye movements, it is necessary for the use to "look" at the target in order to collect the ground truth data. This dependency on the user invokes some inconsistency in the ground truth and significantly decreases the quality of the data. As such, building a high-quality dataset from the ground up and to the scale required to train, test, and validate a reliable eye-tracking system is a challenging proposition. A possible path forward lies in novel approaches that leverage the comparatively easy to acquire 2D eye-images. We posit that sufficient information is available in the 2D eye images to infer gaze and to synthesize eye images for any given user.

The goal of this workshop is to raise awareness of the new challenges that VR and AR present to the eye tracking community, and to engage the broader computer vision and machine learning community in a discussion around these challenges. This workshop will host a competition that is structured around 2D eye-image datasets that we have collected using a prototype VR head mounted device. Entries to this competition will address some outstanding questions relevant to

the application of eye-tracking for VR and AR platforms. We anticipate that the dataset released as part of the competition will also serve as a benchmark dataset for future research in eye-tracking for VR and AR.

Below is the list of topics that are of particular interest for this workshop:

- Semi-supervised semantic segmentation of eye regions
- Photorealistic reconstruction and rendering of eye images
- Generative models for eye image synthesis and gaze estimation
- Transfer learning for eye tracking from simulation data to real data
- Eye feature encoding for user calibration
- Temporal models for gaze estimation
- Image-based gaze classification
- headset slippage correction, eye-relief estimation
- Realistic avatar gazes

1.1. Submission Details

In addition to the submissions addressing competition tasks described in Section 1.2, we will have an open submission format where researchers can share their work and novel research ideas for eye-tracking technology. Submission of research works of particular interest to the workshop are encouraged. The submitted papers must follow ICCV paper submission guidelines. Submissions will be evaluated by the program committee and selected papers will be published in ICCV workshop proceedings.

1.2. Challenge Details

We plan to host two challenge tasks as part of the workshop. For each of these challenges, we will host a website https://openeds.org, where the ground-truth data and baseline results for each task will be made available. Participation in the challenges will require the submission of a workshop paper, following ICCV paper submission guidelines, detailing the algorithm and proposed solution to the corresponding challenge task.

The top three-winning solutions in each category will receive a cash prize sponsored by Facebook as listed in Section 3.5. The winner of each challenge will also be invited for a oral presentation of their work. A subset of the challenge submissions will also be considered for poster presentations.

We will host an evaluation server, where each participant will submit their results on the test set in the specified format. The platform will then automatically run the evaluation script on the results and place the scores in the leaderboard. The ground truth for the test dataset will remain hidden from the challenge participants and will be used to score model accuracy per the proposed evaluation criterion for each challenge.

We ask each participant to upload one result from their best algorithm. The participants may choose to report results in their papers for multiple versions of their algorithm (e.g., with different parameters or features).

Finally, we will not impose any restrictions for authors to upload their code to open-source code sharing platforms such as github.

The two challenge tasks are:

a. **Semantic segmentation challenge to detect key eye-regions**
A vision-based eye tracking system usually includes two stages, namely, *eye detection* and *gaze tracking*. Eye detection involves localizing eye regions (pupil and iris) in the image, while gaze tracking involves the estimation of 3D gaze co-ordinates. This challenge is focused on the *eye detection* stage, with the goal of designing a machine learning solution to detect and segment the three-key eye regions, i.e., sclera, iris and pupil from an input 2D eye-image.
The participants will have access to training and validation datasets comprising of pairs of ground truth semantic labels and the corresponding 2D eye-images. The challenge seeks models that generates a semantic-mask for a given 2D eye-image, labelling the 3 distinct eye-regions: sclera, the pupil and the iris.

The winning entry will be evaluated using the performance metric ($0 < M \leq 100$), defined as follows:

$$M = 50(P + \min(\frac{1}{S}, 1))$$

where, $0 < P \leq 1$, measures model-accuracy as defined by the unweighted mean intersection-over union score over all classes for the test-set, and $S > 0$, measures model-complexity, as defined by the number of model parameters, measured in unit of model-size in MB. More precisely, model-size in MB will be computed as follows: S= number of learned model parameters x 32 (floating point) / (1024*1024).

The choice of metric M is motivated by the fact that we want to encourage solutions that offer the best trade-off between model performance and model complexity. We impose a lower bound of 1 MB on model complexity such that for all models that meet the complexity constraint of $S \leq 1$ MB, model-accuracy takes precedence. The 1MB model size constraint is derived from hardware limitations for real-time inference using these models in an end-to-end eye-tracking pipeline on VR and AR platforms.

b. Synthesize realistic eye-images

The dynamics of the periocular regions are an important contributor to the estimation of the point of gaze in machine learning-based VOG systems. However, specific mechanisms of this contribution are unknown today to a large extent. Thus, a system that is able to generate synthetic images with realistic periocular regions at different eye rotations would aid in enabling the exploration of novel techniques for high quality eye tracking.

For the second challenge, we require participants to synthesize eye images, given a semantic segmentation mask and a set of eye images specific to a user. The training images are captured in the same session and have the same environmental parameters, such as illumination, camera pose. The test set will include a diverse set of segmentation masks, and we expect the participants to submit one synthetic eye image per mask. The synthetic images will then be compared with the actual eye images that belong to the same user and which root the same segmentation mask. Submissions will be scored based on image space metrics.

Dataset Specifications:

We will release a dataset of images for 150 participants, recorded with a head-mounted display under infrared illumination. This data is recorded at 200Hz and accompanied by information on calibration of the lights and cameras.
In addition to around 20,000 images with semantic segmentation annotations, we will release two additional sets of unlabeled eye images: (a) a random set of 60,000 eye-images and (b) 4s long video sequences of eye images, roughly totaling 60,000 images. We ensure that the whole dataset includes a variety of difficult cases and biometrical variety among the subjects. A python API for loading and evaluating this data will be provided. We plan to release full details of the dataset on the competition website at https://openeds.org.

1.3. Relation to Previous Workshops

- ACM Symposium on Eye Tracking Research and Applications (ETRA) 2018

Facebook Reality Labs also organized a workshop at ETRA 2018, a premium conference for eye tracking research and applications, focusing on the specific challenges for eye tracking in a VR and AR environment. The general theme of ETRA 2018 was well aligned with our goal of enabling pervasive eye tracking for everyone. However, much of the research presented at the conference was focused on applying current approaches and working around their limitations, rather than advancing the state of the art or investigating novel approaches to determining the eye state. This highlights the importance of involving the larger scientific community outside of research groups focusing on eye tracking specifically.

- IEEE Workshop on Biometrics – CVPR 2018

This workshop aims to improve research towards information extraction from biometrics. Even though the main motivation leans towards security applications, the ultimate goal of increasing the coverage of the population that the technology can reach is the same. Our workshop can be thought as an overlap with biometrics research applicable to VR. Also, both workshops aim to speed up and balance the research done at academic and private research organizations as well as government labs.

- Computer Vision for VR – ICCV 2017

Although this workshop explored the broader topic of 3D vision and graphics research for virtual reality, some sessions in this workshop are related to our proposed workshop about improving the current state-of-the-art in human-oriented interaction techniques in VR environments from a hardware and research perspective. In particular, the workshop included papers in eye/head tracking for improving VR/AR headsets, which overlap with the main motivation for our proposal.

- Perception-driven Graphics and Displays & Eye Tracking and Vision Augmentation – IEEE VR 2019

Both of these workshops cover specific advances in perception and eye tracking in VR, with the motivation of discovering novelties in human perceptual system and eye tracking solutions for more immersive and comprehensible experiences in VR. We would like to carry a similar

mission; however, we would like to merge these communities with computer vision and machine learning community to enable the development of novel machine learning and computer vision approaches from these collaborations. As a bonus, our workshop also holds a challenge and releases a dataset to facilitate this process.

2. Organizers and Speakers

2.1. Organizers, background and experience *(listed in alphabetic order)*

- Robert Cavin, Facebook Reality Labs, robert.cavin@oculus.com: Robert leads Eye Tracking Research at FRL. Robert received a Masters Degree in Computer and Electrical Engineering from the University of Florida in 2001, with a focus on Robotics, Machine Learning, and Computer Architecture. He was a Senior Computer Architect at Intel Corporation from 2001 to 2010, and a founding member of the Larrabee Massively Mulitscalar x86 architecture team that led early work on GPGPU accelerators. After leaving Intel, Robert founded a mobile app-based business and sold this in early 2012. Robert joined Oculus Research in 2014. He identified the need for eye tracking beyond the state of the art, and built a team of top researchers in Computer Vision, Machine Learning, and Optics to deliver eye tracking solutions that unlock the vision for VR and AR.

- Jixu Chen, Computer Vision Engineer, Facebook jixu.chen@Oculus.com: Dr. Jixu Chen is a computer vision engineer at Facebook. He received the Ph.D degree in Electrical Engineering from Rensselaer Polytechnic Institute (RPI), Troy. NY in 2011, with specialization on human behavior tracking with computer vision techniques. Dr. Chen's work has been focused on real-time eye tracking and face analysis systems, and their applications in industry. From 2011 to 2015, he is the lead research scientist at GE Global Research, working on a multi-person social interaction analysis system. Before joining Facebook AR/VR team, he has spent 2 years working on eye tracking and computer vision algorithms with several AR startup companies. Jixu has published over 20 peer-reviewed publications in top computer vision journals and conferences and 3 book-chapters on eye tracking and face analysis

- Ilke Demir, Deep Learning Research Scientist, DeepScale idemir@purdue.edu Ilke Demir earned her PhD Degree in Computer Science from Purdue University in 2016, focusing on 3D vision approaches for generative models, urban reconstruction and modeling, and computational geometry for synthesis and fabrication. Afterwards, she joined Facebook as a Postdoctoral Research Scientist working with Ramesh Raskar. Her research included human behavior analysis and deep learning approaches in virtual reality, geospatial machine learning, and 3D reconstruction at scale. In addition to her publications in top-tier venues, she has organized several workshops and competitions (WiCV, DeepGlobe, SUMO, SkelNetOn) in the intersection of deep learning and computer vision, including the renowned DeepGlobe workshop at CVPR18. She has also been involved in organizing committees of several events and workshops (in CVPR, SIGGRAPH, ACCV, ECCV, NeurIPS, etc.).

- Stephan Garbin, PhD Student, University College London, Microsoft Research, Facebook Reality Labs, stephangarbin@outlook.com: Stephan is currently a PhD candidate working

on vision and machine learning at University College London under the supervision of Professor Gabriel Brostow and Dr Jamie Shotton.

- Oleg Komogortsev, Visiting Scientist/Associate Professor, Facebook Reality Lab/Texas State University, oleg.komogortsev@gmail.com: Dr. Komogortsev is currently a tenured Associate Professor at Texas State University and a Visiting Scientist at Facebook Reality Labs. He has previously worked for such institutions as Johns Hopkins University, Notre Dame University, and Michigan State University. Dr. Komogortsev conducts research in eye tracking with a focus on sensor design, cyber security (biometrics), human computer interaction, usability, bioengineering, and health assessment. This work has thus far yielded more than 100 publications and several patents. Dr. Komogortsev's research was covered by the national media including NBC News, Discovery, Yahoo, Livesience and others. Dr. Komogortsev is a recipient of two Google Virtual Reality Research Awards and a Google Faculty Research Award. Dr. Komogortsev has also won National Science Foundation CAREER award and Presidential Early Career Award for Scientists and Engineers (PECASE) from President Barack Obama on the topic of cybersecurity with the emphasis on eye movement-driven biometrics and health assessment. In the past Dr. Komogortsev co-organized various competitions on eye movement biometrics and served as a competition co-chair at prestigious biometric conferences. Dr. Komogortsev's current grand vision is to push forward eye tracking solutions in the future virtual and augmented reality platforms as enablers of more immersive experiences, security, and assessment of human state.

- Immo Schuetz, Postdoctoral Research Scientist, Facebook Reality Lab, ischtz@Oculus.com
  Immo Schütz is a Postdoctoral Research Scientist with Facebook Reality Labs, where his current research focuses on determining technical and perceptual requirements for different applications of eye tracking technology. Immo graduated with a PhD in Perceptual Psychology from Justus-Liebig University Giessen in 2015. He then spent two years in the Physics of Cognition group at Chemnitz University of Technology, where he developed research methods for eye tracking in visual attention and perceptual awareness research, before moving into industry and joining Facebook Reality Labs in 2017. Immo has extensive experience with eye tracking performance metrics in existing VR solutions and co-organized an FRL sponsor workshop at ETRA 2018.

- Abhishek Sharma, Research Scientist, Facebook Reality Lab, abhisharayiya@gmail.com: Abhishek graduated from the department of Electrical Engineering at the Indian Institute of Technology Roorkee in 2010 and completed his Ph.D. in Computer Science from University of Maryland, College Park in 2015 with a focus on computer vision and machine learning. During his Ph.D., he worked on learning with multiple views of data and developed mathematical models for bringing the canonical knowledge from different views into a common representation for cross-view data matching and fusion. He has worked on face-recognition, object classification/detection, human pose-estimation, semantic segmentation and internet-scale reverse image search. He joined Apple after his Ph.D. and contributed to the foundational development of FaceID matcher for iPhone X. In 2017, he co-founded Gobasco Pvt. Ltd., an agri-tech startup in India that leverages data-analytics and AI served as its Director and CTO before joining Facebook.

- Sachin Talathi, Research Scientist, Facebook Reality Lab. Sachin.talathi@oculus.com: Dr Talathi has an extensive background in academic and corporate research with over 60 peer-reviewed published research paper and 12 patents in topics from Computational Neuroscience, Neural Signal Processing and Machine Learning. Dr Talathi graduated from University of California, San Diego with a PhD in Physics in 2006. His academic research program, as tenure track faculty in the Department of Biomedical Engineering at University of Florida, spanned the gamut of developing novel machine learning solutions to analyze large amounts of biomedical data to developing computational models to explore pathological brain dynamics. During his academic tenure, Dr Talathi co-organized and participated in several conference workshops focusing the role for computational methods to understand brain disease. He left academia in 2013, to pursue a corporate career at Qualcomm Research in San Diego, where his work focused on developing low complexity deep learning models for real-time inference on mobile platform. During this time, Dr Talathi co-organized an ICML workshop on low-complexity deep-learning, as well as published papers on related topics in leading machine learning conferences such as ICML, ICLR and ICIP. Before coming to Facebook Reality Lab, Dr Talathi spend a year and a half at Amazon Inc in Seattle, working on big data and machine learning problems in Amazon Retail. At FRL, Dr Talathi's work is focused on developing Machine Learning strategy to solve the grand challenge of eye-tracking for AR and VR.

2.2. Invited Speakers

- Oleg Komogortsev (Confirmed)
- Satya Mallick (Confirmed)
- Ramesh Raskar (Confirmed)
- Ji Qiang (Invited)
- Wolfgang Fuhl (Invited)
- Michael Stengel (Invited)
- Andreas Bulling (Invited)
- Richard Szeliski (Invited)

3. Logistics

3.1. Format

Full-Day workshop comprising of a series of invited and contributed talks, and a poster session.

3.2. Estimated Participation

We will invite submissions that contain novel work in the categories described above, and current or previously published work which is benchmarked on our dataset. The top two winners for each challenge task will be invited for an oral presentation. In addition, a subset of authors of selected papers will also be invited for an oral presentation. All the accepted papers will be invited for a poster session. We expect around 40 submissions for both the challenge and the workshop, and around 12 papers accepted for oral/poster presentations to be included in the workshop proceedings.

We anticipate participation from an interdisciplinary audience from academia and industry working in the AR/VR space, from eye tracking, biometrics, computer vision, computer graphics, optics, and machine learning communities. We expect around 80-100 attendees in addition to the presenters.

3.3. Tentative Program

The tentative program is as follows:

- 09:00-09:15 Welcome

- 09:15-09:45 Invited Speaker 1
- 09:45-10:00 Oral 1
- 10:00-10:15 Oral 2
- 10:15-10:30 Morning Break

- 10:30-11:00 Invited Speaker 2
- 11:00-11:15 Oral 3
- 11:15-11:30 Oral 4
- 11:30-1:00 Lunch Break

- 1:00-1:30 Invited Speaker 3
- 1.30-1.45 Oral 5
- 1.45- 2.00 Oral 6
- 2.00- 2.15 Afternoon Break

- 2:15-2:45 Invited Speaker 4
- 2.45- 3:00 Oral 7
- 3.15- 3:30 Oral 8

- 3.45-4:45 Poster Session
- 4.45-5.00 PM Award ceremony and closing

3.4. Technical/Program Committee

- Kaan Aksit, NVIDIA (confirmed)
- Andreas Bulling, University of Stuttgart (confirmed)
- Rob Cavin, FRL (confirmed)
- Jixu Chen, FRL (confirmed)
- Ilke Demir (confirmed)
- David Dunn, UNC (confirmed)
- Oleg Komogortsev, Texas State University, FRL (confirmed)
- Ramesh Raskar, MIT (confirmed)
- Immo Schuetz, FRL (confirmed)
- Sachin Talathi, FRL (confirmed)
- Lei Xiao, FRL (confirmed)
- Marina Zannoli, FRL (confirmed)

- Rachel Albert, NVIDIA (tentative)
- Jennifer Gille, Oculus (tentative)
- Xia Zhou, Dartmouth (tentative)

3.5. Financial Support

The workshop and the competition are funded by Facebook. Facebook will offer financial support to cover ICCV registration fees as well as all travel and accommodation expenses for all invited speakers and for the first author of the first place winner for each of the two challenge tasks.

In addition, Facebook will sponsor the following prices for the top-3 winners of the two-challenge tasks: The first-place winning entry for each competition will receive prize money of $5000, the second-place winning entry will receive prize money of $3000 and the third-place winning entry will receive prize money of $2000.

4. Important Dates

- Call for participation and dataset release: April 25, 2019
- Competition ends and submissions close: August 10, 2019
- Notifications to authors: August 25, 2019
- Camera ready deadline: August 30, 2019
- Workshop: October 27 or November 2nd, 2019