Corneal Reflection Detection and Matching Using Deep Learning for Eye Tracking

Soumil Chugh, Braiden Brousseau, Prof Jonathan Rose & Prof Moshe Eizenman

Electrical and Computer Engineering Department, University of Toronto

UNIVERSITY OF TORONTO

Introduction

.

What is Eye Tracking?
Ability to determine where a person is looking on a display or in the real world.

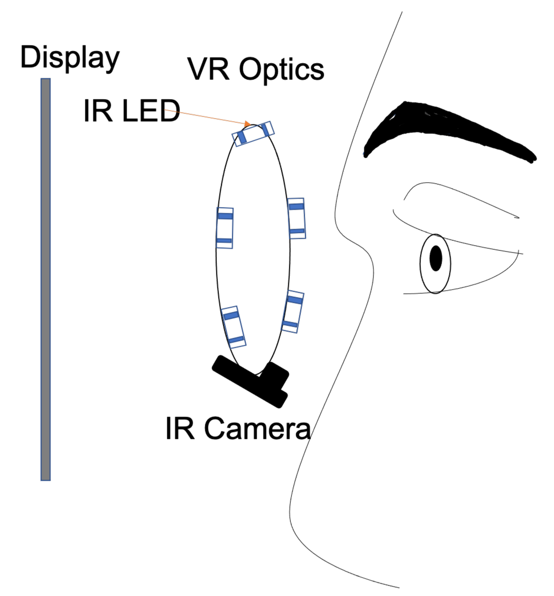

Fig 1. Eye Tracking in virtual reality (VR) Systems

Method of detecting gaze [1]
Model based approach: Uses two or more corneal reflections and pupil center to estimate the point-of-gaze. These estimates are insensitive to headset shifts.

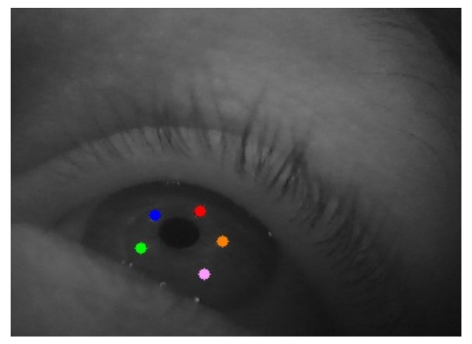

Fig 1. Labelled Corneal Reflections

Corneal reflection Requirements
- Precise locations of corneal reflections in eye images (Detection).
- Information about which reflection came from which light source (Matching)

Challenges:

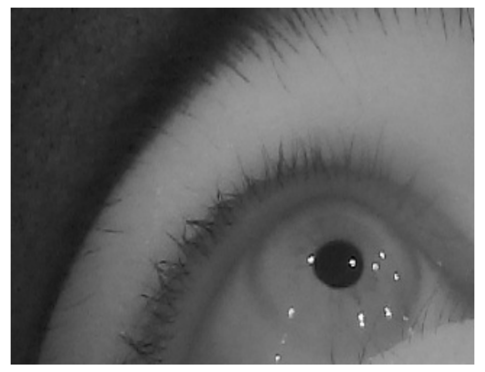

Fig 2: Natural Spurious Reflections

Fig 4: Missing Reflections with change in gaze angle

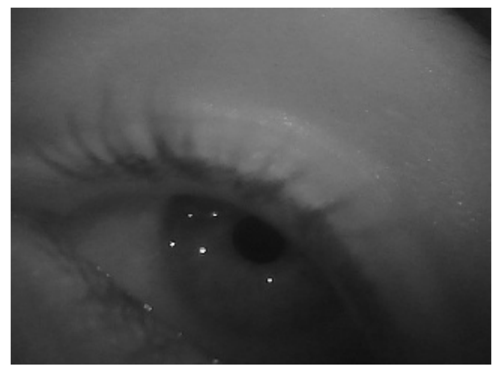

Fig 3: Missing reflections due to eyelid occlusion

Methods

Convolutional Neural Network
- U-Net style architecture
- Input and Output share the same size
- Loss Function: Weighted combination of Cross Entropy and Soft Dice Loss

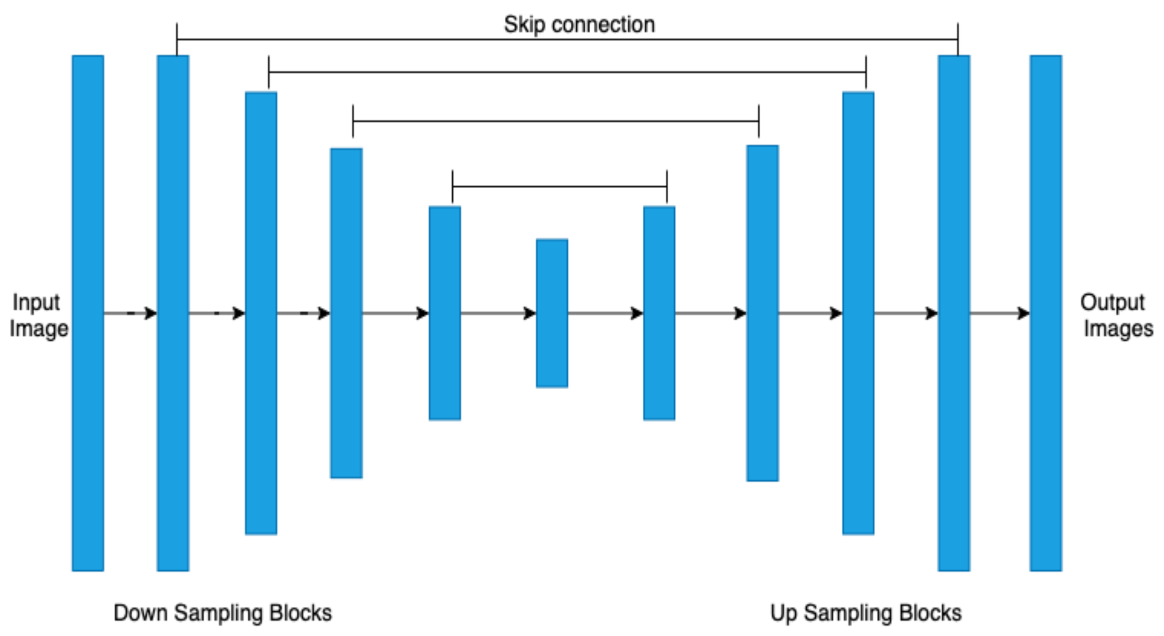

Fig 5. UNET Architecture

Dataset Collection and its Labelling
- Eye images (640x480) collected from 15 people inside an HTC Vive VR headset.
- Three of the 15 subjects wore glasses during the data collection.
- Also used unlabeled dataset from NVIDIA [3] which consists of eye images of 10 people.
- Manually labelled 4000 eye images across 25 people.

Static Data Augmentation
- Increased dataset from 4000 to 40000 eye images
- Generated 10 random crops of resolution 320x240 for every labelled eye image of resolution (640x480).

Dynamic Data Augmentation
- Added Spurious Reflections with varying intensity and sizes.
- Added Gaussian and Motion Noise.
- Adjusted Brightness and Contrast of image.

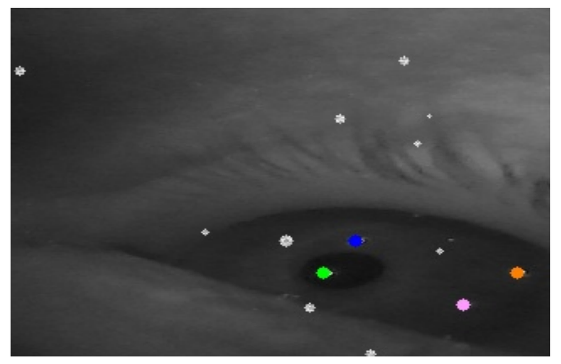

Fig 6: With Spurious Reflections only

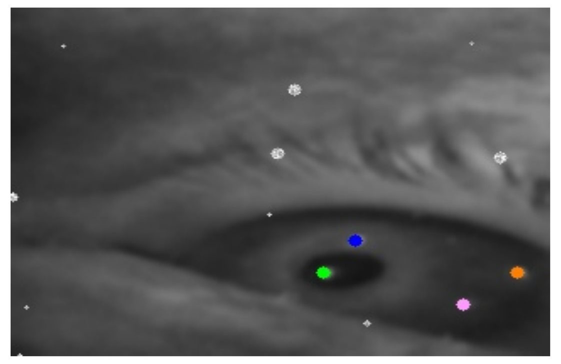

Fig 7: All techniques applied

Results

Dataset Splitting
- Dataset was divided into training (76%), validation (12%) and test (12%) set.
- Each dataset contained images from different subjects.

Single Corneal Reflection Performance
- Accuracy for detecting true corneal reflections and finding its matching light source on the test dataset is 91%.
- One forward pass through the neural network on a RTX2060 GPU takes 5.7ms.
- Uses 3 Mb of memory.

Performance Comparison with other Systems
- Only one prior deep learning approach exists for eye tracking in VR systems [2].
- Accuracy comparable with [2].
- 10x Faster and uses 33x less space in memory than [2].

Effect of Spurious Reflections

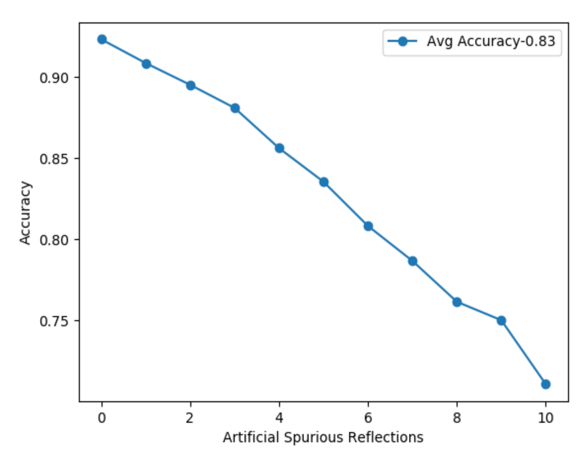

Fig 8: Corneal reflection pair accuracy

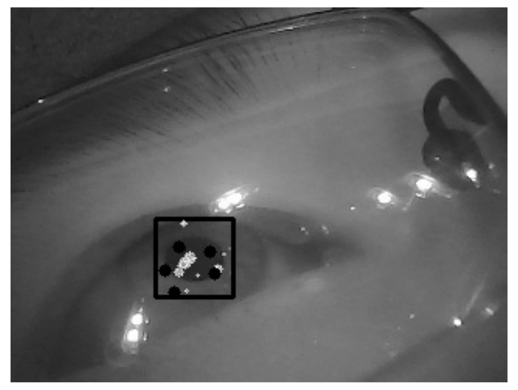

Fig 9: Addition of spurious reflections near the pupil

Eye Tracking Performance
- Integrated the trained network in a VR eye tracking system.
- Made use of a similar UNET style network for pupil estimation.
- Tested the eye tracker on 3 subjects.
- System reported mean absolute error of 1°

Table 1: Eye Tracking Performance

Metric	Gaze Error (°)		
	EyeNet [2]	NvGaze [3]	Our System
Mean Absolute Error	3	2.1	1

Conclusions

- Presented a neural network based on U-Net architecture for tracking multiple corneal reflection inside a VR system.
- The only prior deep learning-based solution [2], uses a series of networks for detecting and matching corneal reflections.
- Use of U-Net style network allows training with a dataset 24x smaller compared to [2].
- The proposed network is computationally efficient than [2], while achieves similar accuracy.
- Under the presence of upto 10 spurious reflections near the pupil, the accuracy of tracking multiple corneal reflections only reduces by 8%.
- Easy to retrain a U-Net style network for different hardware configurations.
- Performance of the network is improved with the use of Dilated Convolutions in the CNN layers and max pooling layer for down sampling.
- The proposed algorithm when integrated into an eye tracking system within a VR headset achieves an error of 1°.
- The gaze error is at least 100% lower than previous VR eye tracking systems.
- The presented approach is applicable to other XR systems that use Augmented Reality and Mixed Reality headsets.

Bibliography

1) E. D. Guestrin and M. Eizenman, "General theory of remote gaze estimation using the pupil center and corneal reflections," in IEEE Transactions on Biomedical Engineering, vol. 53, no. 6, pp. 1124-1133, June 2006.

2) Z.Wu, S.Rajendran, T.Van As, V.Badrinarayanan and A.Rabinovich, "EyeNet: A Multi-Task Deep Network for Off-Axis Eye Gaze Estimation," In 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), pp. 3683-3687, October 2019.

3) J. Kim, M. Stengel, A. Majercik, S. De Mello, D. Dunn, S.Laine, M. McGuire, and D. Luebke, "Nvgaze: An anatomically-informed dataset for low-latency, near-eye gaze estimation," InProceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Scotland, pp. 1-12, May 2019.