Gaze Direction Tracking in Infants

Internship Report submitted by

Prashanth Reddy Duggirala
Roll No: CED14I006

Under the Guidance of
Koteswararao Chilakala
Center For Innovation,
LV Prasad Eye Institute

in partial fulfilment for the award of the degree

Dual Degree
in
Computer Engineering

Indian Institute of Information Technology Design and Manufacturing
Kancheepuram, Melakottaiyur, Chennai-600127

October 2018

Contents

List of Figures

Abstract

Viewed in the context of machine vision, successful gaze tracking requires techniques to handle imprecise data, noisy images, and a possibly infinitely large image set. The most accurate gaze tracking has come from intrusive systems which either require the subject to keep their head stable, through chin rests etc., or systems which require the user to wear cumbersome equipment, ranging from special contact lenses to a camera placed on the user's head to monitor the eye. The subject here is not an adult, but an infant, wearing any special equipment or keeping the head still is not easily possible, therefore, the system described here attempts non-intrusive gaze tracking,

1 Introduction

Pediatric Perimeter is a novel, first of its kind device to measure the side vision of infants. The procedure of quantifying visual field is called perimetry. This device is used on children with potential disease to cause a visual field defects such as Glaucoma, Retinitis Pigmentosa, Retinopathy of Prematurity, etc. It involves checking the eye/head movement of an infant to a visual stimulus in particular areas in the visual field using various tests in perimetry involving LEDs fitted inside the Pediatric Perimeter [1]. The project's aim is to automatically acknowledge a child's reaction to a visual stimuli provided in the Pediatric Perimeter device. The device is a hemispherical dome with light emitting diodes (LEDs). The LEDs are controlled using a computer program to measure Reaction Time (RT) to Gross Visual Fields (GVF) and the Visual Field Extent (VFE). Infants lie down in the dome positioned in a dark room. Eye or head movement towards the stimuli is monitored with an infrared (IR) camera.

Pediatric Perimeter is currently operated by a trained clinician, who has to acknowledge the reaction from the children during testing inside a Pediatric perimeter. This approach has an inherent issue of human errors and subjective biases in quantifying Reaction Times and judging direction of eye gaze. The idea is to incorporate gaze detection in the device's software which aids the user in performing the tests with reduced human error and help in getting more accurate readings.

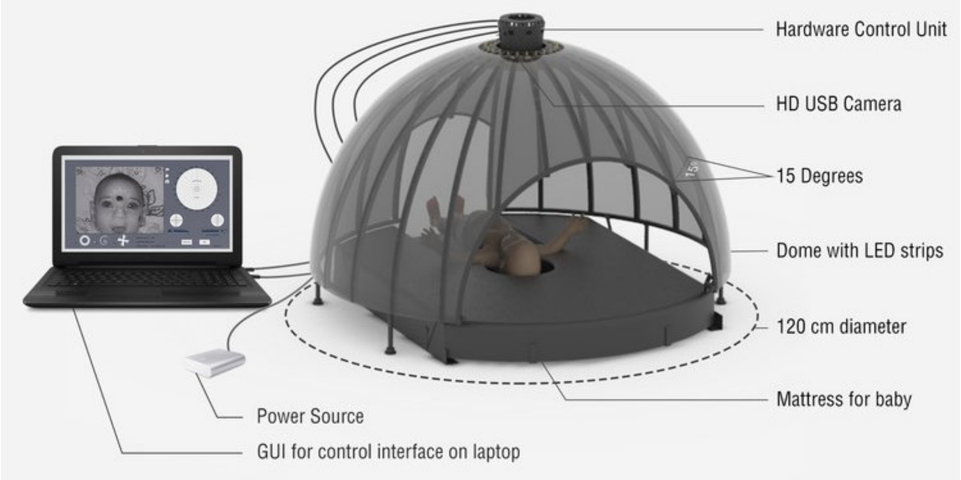


Figure 1: **Pediatric Perimeter**

Source: LVPEI Center For Innovation

2 Preliminaries

Most modern approaches to remote, non-contact point-of-gaze estimation are based on the analysis of eye features extracted from video images. The most common features are the centers of the pupil and one or more corneal reflections. The corneal reflections (first Purkinje images) are virtual images of infrared light sources that illuminate the eye, and are created by the front surface of the cornea, which acts as a convex mirror. The first Purkinje image off the corneal surface is also called the glint. Typically, point-of-gaze estimation systems have to be calibrated for each subject by having the subject fixate on multiple points in the scene. A multiple-point calibration procedure, however, can be an obstacle in applications requiring minimal subject cooperation such as applications with infants [2]

In standard gaze trackers, the image of the eye is processed in three basic steps. First, the specular reflection of a stationary light source is found in the eye's image. Second, the pupil's center is found. Finally, the relative position of the light's reflection and the pupil's center is calculated [3]. From information about their relative positions, the gaze direction is determined (Figure 2).

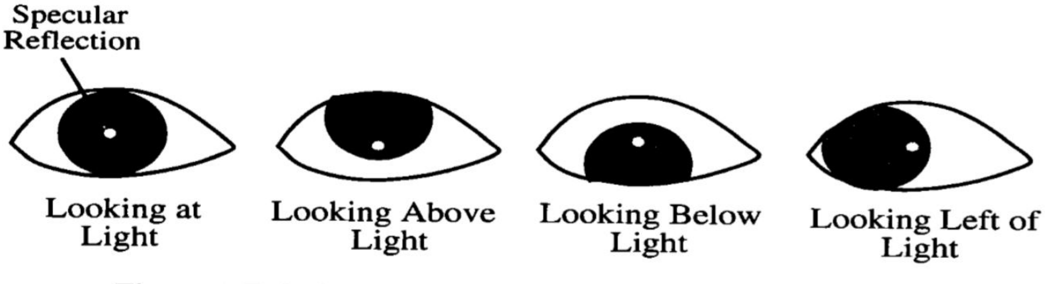


Figure 2: **Relative position of specular reflection and pupil.**

Source: Dean Pomerleau & Shumeet Baluja, School of Computer Science, Carnegie Mellon University

The first and fourth Purkinje images can be used for tracking the direction of gaze by the Dual-Purkinje Image technique [4], which uses the relative positions of these reflections to calculate the direction. The Dual-Purkinje-Image technique is generally more accurate than the prior technique, but the main disadvantage with this technique is that the fourth Purkinje image is rather weak, so the surrounding lighting must be heavily controlled.

Another system uses the PupilCenter/Corneal-Reflection (PCCR) method to non-intrusively determine the eye's gaze direction. The main concept of the PCCR technique is to locate the pupil center and the center of the corneal reflection off the eye surface, and use these two centroids to determine the gaze direction of a user. To this end, a video camera is oriented close to the user's nominal gaze direction such that it is focused on the user's eye [5].

Naqvi et al. [6] proposed a corneal reflection-based gaze-tracking system for assisted driving. They installed an infrared camera in the vicinity of the car dashboard, and used Dlib [7] with three CNNs to extract the left eye, right eye, and face features to estimate the participant's gaze. Choi et al. [8] detected driver faces with a Haar feature face detector and used CNN to categorize gaze zones.

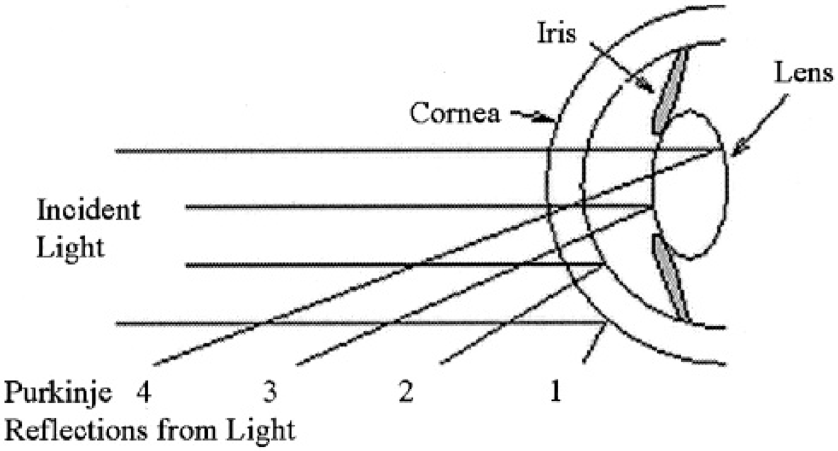


Figure 3: **Purkinje Reflections on the eye**

Source: Optical Society of America

3 Novelty

Since the device is used on infants, The project demands a solution that is:

- Non-Intrusive
- Calibration Free
- Head Pose Invariant

Gaze tracking with only an IR camera was not an area where extensive work has been done. Further, Detecting gaze on infants is more unique, owing to the fact that there are no known open source datasets which contains infant head poses and gaze information. We do not require precision in terms of gaze angle, but rather an accurate detection of the gaze direction of the infant.

The above ideas of gaze tracking in assisted driving have served as an inspiration where the interior of an automobile can be likened to the inside of the Pediatric Perimeter.

4 Contribution

4.1 Tools Used

- Languages: Python, C
- libraries: TensorFlow, Keras, OpenCV, ffmpeg
- Hardware:
  - CPU: Intel i7 7700
  - GPU: Nvidia GeForce GTX 1080Ti

4.2 Available Raw Data

4.2.1 Operating Procedure of the device

The procedure started with testing the gross visual field; Four quadrants (top -right; left, Bottom- right; left), later if the infant cooperated, the cardinal meridians (45°, 135°, 225°, 315°) following which rest all 20 meridians were tested either completely or sequentially/sweep (peripheral LED to central LED).

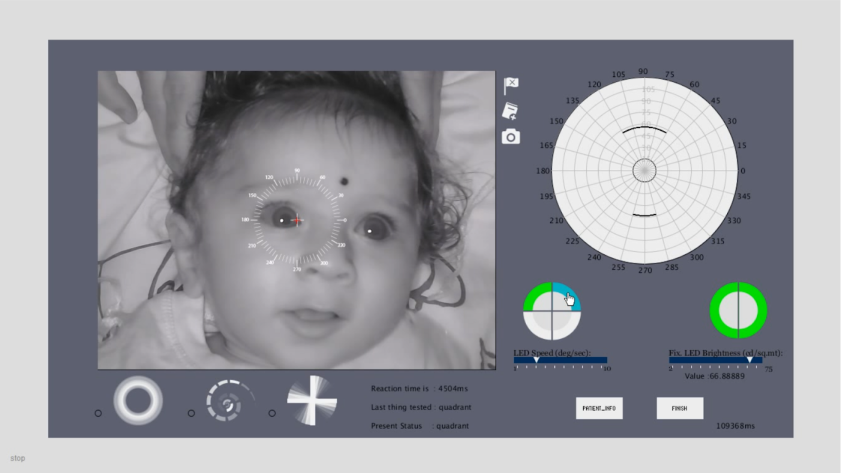
The response was taken as an eye movement towards the stimulus LED. Later, the video generated was analyzed manually by three different optometrists using an open-source software to get the precise reaction time to each stimulus.

4.2.2 Format of the Data

The data we have from the device as mentioned above are the screen recordings of the tests performed and csv files containing the reaction time information in the form of timestamps.

Training data was taken from the tests which had cooperative subjects; For whom all kinds of meridians and sweep tests were performed. This meant that all possible orientations of the eye were available.

The training data was collected from 54 tests and an additional 34 tests were taken for Validation.



Reaction time is : 4504ms
Last thing tested : quadrant
Present Status : quadrant

LED Speed (deg/sec):
Fix. LED Brightness (cd/sq.mt):
Value :66.88889

109368ms

Figure 4: **Typical example of a frame from the videos**

FB_19.onset	FB_19.offset	Test Performed	Status
2201	35241	L&R Hemi	NEFS
61761	66280	TR QUAD	FEFS
70521	70961	TR QUAD	NENS
73261	73782	BL QUAD	NENS
76080	76683	BR QUAD	NENS
87621	88042	TL QUAD	NENS
96041	98543	MERIDIAN 45	NEFS
102081	102541	MERIDIAN 135	NENS
105081	105682	MERIDIAN 225	NENS
107841	108143	MERIDIAN 315	NENS

Table 1: **Typical example of the timestamp information as analysed by optometrists**

Due to the limitations of the available raw data, It was decided to try estimating the 9 gaze directions in which a baby can look were identified and are as follows;

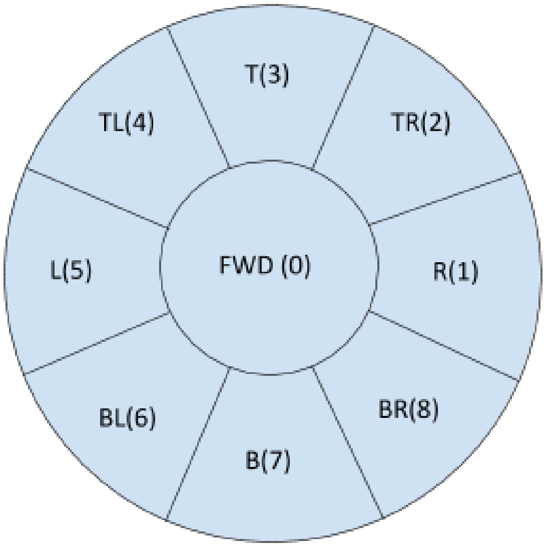


Figure 5: Regions of Gaze along with their class ID

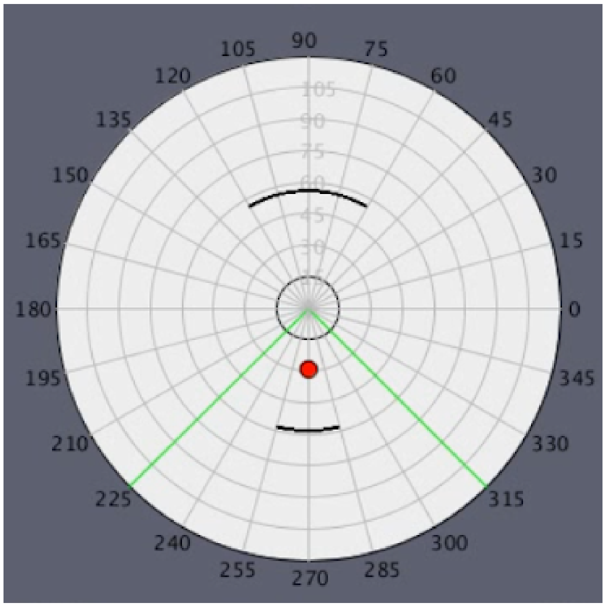


Figure 6: Meridians and their angles

4.3 Data Set Generation

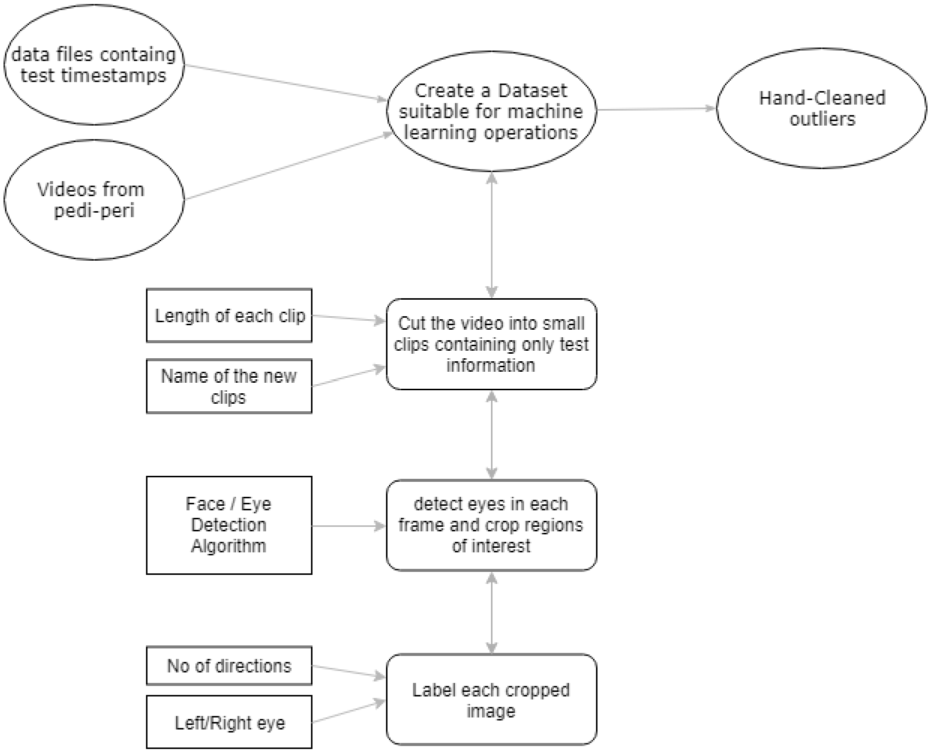

data files containg test timestamps
Videos from pedi-peri
Create a Dataset suitable for machine learning operations
Hand-Cleaned outliers
Length of each clip
Name of the new clips
Cut the video into small clips containing only test information
Face / Eye Detection Algorithm
detect eyes in each frame and crop regions of interest
No of directions
Left/Right eye
Label each cropped image

Figure 7: **Flowchart Illustrating Dataset Generation Process**

4.3.1 Region of Interest segmentation

Each frame is captured from the clips and then CV - image processing is used to segment the regions of interest with the help of dlib facial landmark detector.

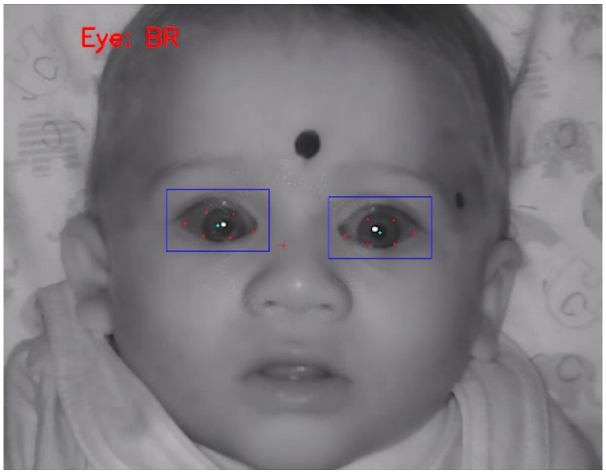


Figure 8: Illustration of a face

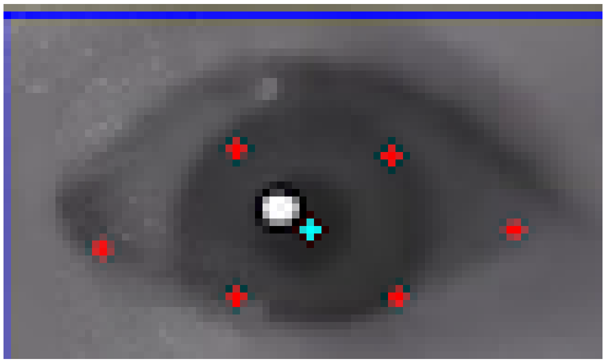


Figure 9: Segmented eye image (Right)

The centroid of landmark points (cyan) marking the eye is used as the reference to denote a standard region to be segmented of size 100 X 60 pixels.

The isopter overlay on the video also has been removed using image processing.

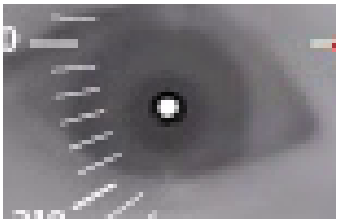


Figure 10: Illustration of the eye with the overlay on top

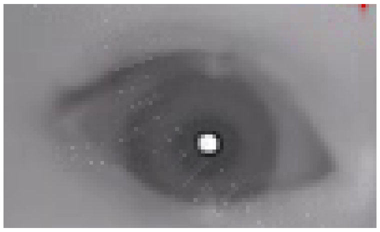


Figure 11: Image of the eye after clearing the overlay

4.3.2 Test Set and Validation Set For each Eye

The Total Number of images taken from videos meant for training and validation data were in upwards of 25,000 and 3,500 respectively. Their Distribution is illustrated as follows;

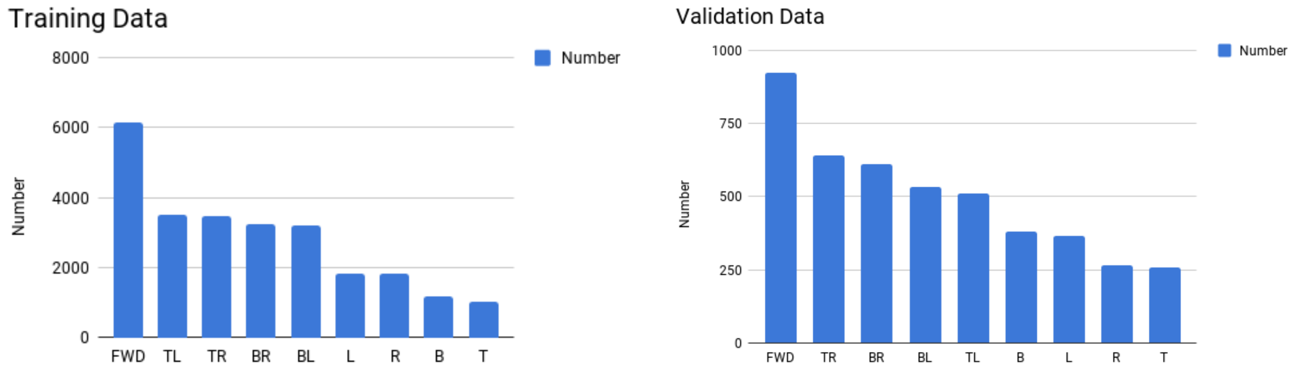


Figure 12: **Distribution of data among the classes**

In the training set, to maintain equal representation of all the classes, the data is sampled to contain 1000 data points from each class.

4.4 CNN For Eye Gaze Direction Classification

If a dataset represents some very specific domain, say for example, medical images or Chinese handwritten characters, and that no pre-trained networks on such domain can be found, then training the network from scratch is the apposite approach.

However, a network's bare-bones architecture can be utilized; Many popular Imagenet [9] Architectures like Inception and ResNets are available for this purpose, But MobileNet [10] is used keeping in mind;

- Small size of the input image
- Smaller model size
- Faster Prediction Time
- Comparable Accuracy

Model	Size	Top-1 Accuracy	Top-5 Accuracy	Parameters	Depth	Input Size
ResNet50	99MB	0.749	0.921	25,636,712	168	224X224
InceptionV3	92MB	0.779	0.937	23,851,784	159	299X299
InceptionResNetV2	215MB	0.803	0.953	55,873,736	572	299X299
MobileNet	16MB	0.704	0.895	4,253,864	88	224X224

Table 2: **Comparing Various State-of-the-art architectures**

Source: Francois Chollet, keras.io

4.4.1 Gaze prediction pipeline

For any instance, a separate model runs on each eye and gives an output vector. Addition of these two vectors gives a resultant, which is used to estimate the gaze direction. This approach is more robust and accounts for peculiar cases like when the baby is not properly aligned to the center of the device.

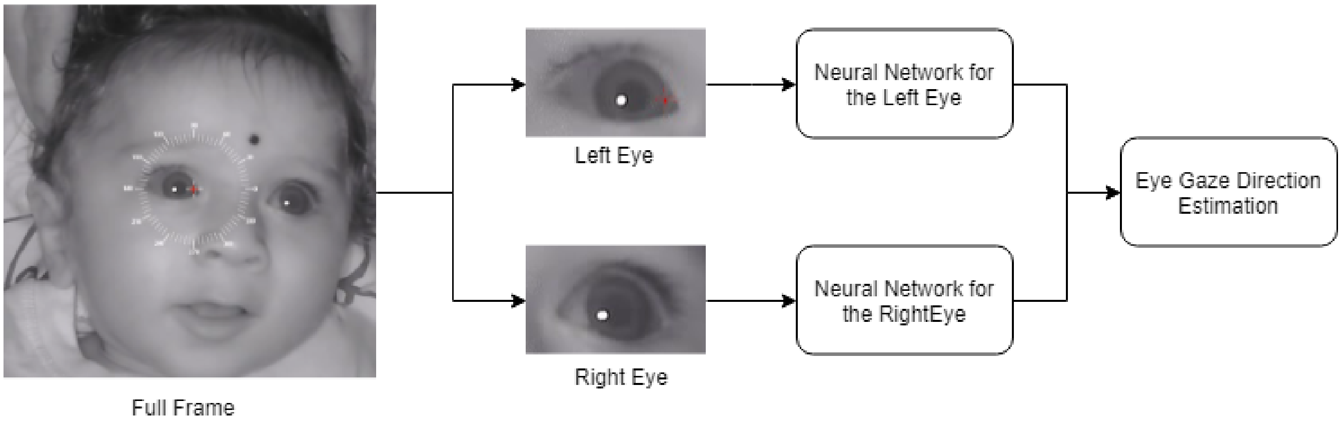

Full Frame
Left Eye
Right Eye
Neural Network for the Left Eye
Neural Network for the RightEye
Eye Gaze Direction Estimation

Figure 13: **Illustrating the prediction process**

4.4.2 Results

The information about the performance of the machine learning model can be understood by a confusion matrix. In general, in a confusion matrix, the predicted classes are compared with the actual classes. Each row of the matrix represents the results of prediction for the corresponding class at that row, while each column represents the actual class. The diagonal cells show the percentage of correct classifications by the trained classifier, while the off diagonal cells represent the misclassified predictions.

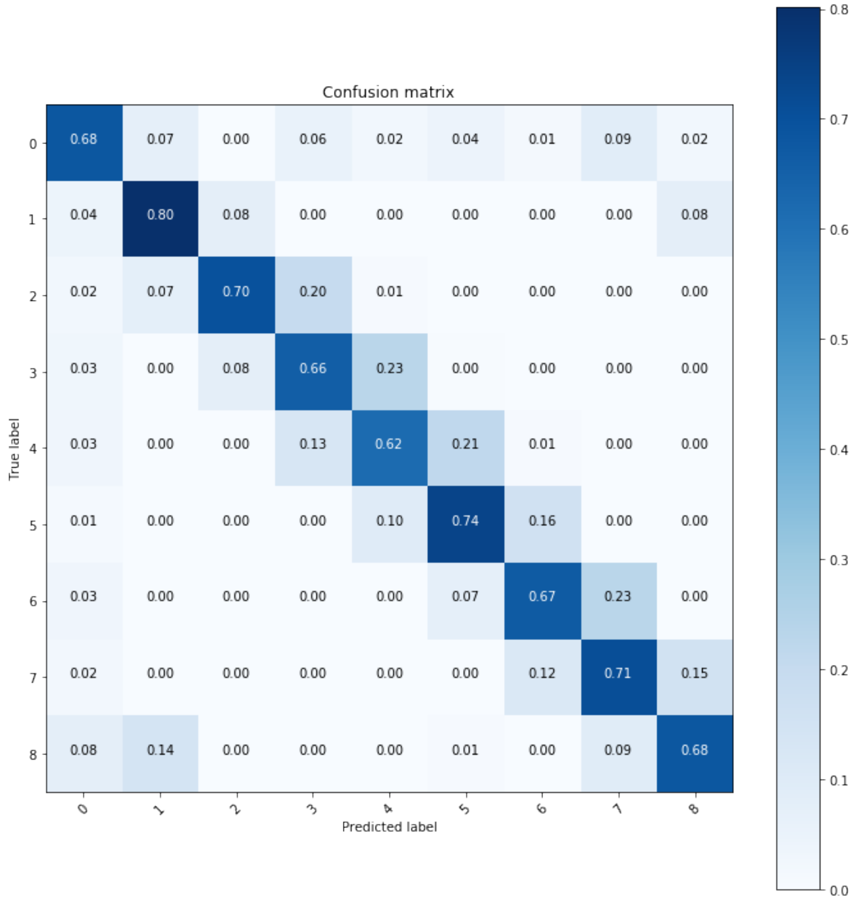


Figure 14: **Confusion Matrix**

From the confusion matrix we can observe that for every True Label, the predictions were distributed slightly to the classes adjacent to them. For Example; Consider Class 5 (L), the predictions were distributed as 74% for the true label, 10% for class 4 (TL) and 16% for class 6 (BL)

This accounts for the cases where the subject is looking at the border of two regions; for example, in between B and BL; then the prediction would be either B or BL, this leads to some ambiguity.

5 Future Work

The course of Research and Development for the pediatric perimeter is detailed below

5.1 3D Head Pose Estimation

Head pose estimation along with Eye Gaze Direction can open new possibilities of utilizing the device for determining visual field extent in infants with conditions like Nystagmus, Strabismus and Cortical Visual Impairment.

5.2 Integration with Device Software

A new version (v4.0) of the Pediatric Perimeter desktop application is being developed on Unity in C. Integrating the Gaze prediction module into this application enables further validation of the models and moreover, helps the module to increase it's learning as more data is gathered, improving the accuracy of the models. The current python scripts have been converted to C, work remains to compile them into a DLL, without any dependencies.

5.3 Stereo Camera Systems

The current camera is a single - 640X480 IR camera, The plan is to device a higher resolution - multiple camera system for capturing more micro-expression information for predicting various conditions like Autism, Cerebral Palsy and Attention-deficit/hyperactivity disorder (ADHD).

References

[1] P. Satgunam, S. Datta, K. Chillakala, K. R. Bobbili, and D. Joshi, "Pediatric perimeter—a novel device to measure visual fields in infants and patients with special needs," *Translational Vision Science Technology*, vol. 6, no. 4, p. 3, 2017.

[2] E. D. Guestrin and M. Eizenman, "Remote point-of-gaze estimation requiring a single-point calibration for applications with infants," in *Proceedings of the Eye Tracking Research & Application Symposium, ETRA 2008, Savannah, Georgia, USA, March 26-28, 2008*, pp. 267–274, 2008.

[3] S. Baluja and D. Pomerleau, "Non-intrusive gaze tracking using artificial neural networks," in *Advances in Neural Information Processing Systems 6* (J. D. Cowan, G. Tesauro, and J. Alspector, eds.), pp. 753–760, Morgan-Kaufmann, 1994.

[4] P. U. Müller, D. Cavegn, G. d'Ydewalle, and R. Groner, "A comparison of a new limbus tracker, corneal reflection technique, purkinje eye tracking and electro-oculography," in *Studies in visual information processing, Vol. 4. Perception and cognition: Advances in eye movement research. Amsterdam, Netherlands* (G. d'Ydewalle and J. V. Rensbergen, eds.), pp. 393–401, North-Holland/Elsevier Science Publishers, 1993.

[5] A. Talukder, J.-M. Morookian, S. Monacos, R. Lam, C. Lebaw, and A. Bond, "Real-time non-invasive eye tracking and gaze-point determination for human-computer interaction and biomedicine," in *International Society for Optical Engineering (SPIE) Defense and Security Symposium, Optical Pattern Recognition XV; April 12-16, 2004; Orlando, FL; United States*, NASA Jet Propulsion Lab., California Inst. of Tech., 2004.

[6] R. A. Naqvi, M. Arsalan, G. Batchuluun, H. S. Yoon, and K. R. Park, "Deep learning-based gaze detection system for automobile drivers using a NIR camera sensor," *Sensors*, vol. 18, no. 2, p. 456, 2018.

[7] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," in *2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014, Columbus, OH, USA, June 23-28, 2014*, pp. 1867–1874, 2014.

[8] I. Choi, S. K. Hong, and Y. Kim, "Real-time categorization of driver's gaze zone using the deep learning techniques," in *2016 International Conference on Big Data and Smart Computing, BigComp 2016, Hong Kong, China, January 18-20, 2016*, pp. 143–148, 2016.

[9] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211–252, 2015.

[10] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *CoRR*, vol. abs/1704.04861, 2017.