Eye prescription verification system

Aditya Sahu
IIIT Delhi

Palak Bansal
IIIT Delhi

Debarshi Dasgupta
IIIT Delhi

Abstract

In todays time we are able to buy lens for eyes online if we have the prescription of our eyes available with us. To obtain the prescription we must physically visit an ophthalmologist [1]. In this paper we attempt to determine the power of lens that a user needs with help of a stereo/mono webcam.

1. Introduction

1.1. Problem Statement

The problem that we aim to solve here is to determine the power of both eyes as accurately as possible. For this after obtaining distances from our depth model we will use physic models to determine each eyes power.

1.2. Input and Output format

The input to our problem will be user's image/video from webcam and output will we two numbers corresponding to left and right eye.

1.3. Challenges involved

One of the biggest challenge involved in our project is to determine the exact distance of eye from the source(webcam). We will take help of depth detection models and geometry to determine the location of eye.

1.4. Motivation

The reason behind picking this project is to be able to convert the process of going to ophthalmologist for an eye checkup online and making the process cost free and fast for everyone.This project can add extra feature for better user experience in apps like Lenskart which aims at seamless user experience for their eyewear in India. The app can add a verification feature by our project where the user can verify the power of the lens they are planning to order on the app. This project can further be extended to verify the doctor's prescription the user is adding to order for the eye glasses. Our project is nowhere planning to replace the medical science testing of eye power but only acts as a verification for the user so that the at home experience through the app is more enjoyable for users.

1.5. Contributions from our side

The major contribution in this project that we will be making is to determine the power of each eye as accurate as possible. This area of research is still very new and not much work has been done in this field. We aim to atleast create an online verification system for eye prescription at the starting stage.

2. Summary of our proposed algorithm

We will be following this procedure to obtain our desired result.

- Capture RGB Image/Video from the webcam which has user's eyes in it.
- Apply de-noise filters like gaussian blur and resize the image for making it ready for next NN layers.
- Detect depth of eye from webcam using Encoder-decoder models.
- Use a physics based model to determine power of each eye.

Main task in our algorithm would be to determine the webcam to eye distance.

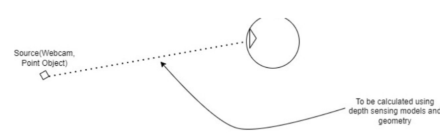

Figure 1. Source to Eye distance

3. Literature Survey

The first step in the process of determining the power of each eye would be to determine the distance of eye from the webcam at which user is able to see a specific font size. For this we would be using depth sensing models. Depth is extracted from either monocular(single) or stereo(multiple views of a scene) images. Traditional methods use multi-view geometry to find the relationship between the images. Newer methods can directly estimate depth by minimizing regression loss [5] or by learning to generate a novel view from a sequence.There has been great advancement in self-supervised depth estimation [3]. which is particularly exciting and groundbreaking! In this method, a model is trained to predict depth by means of optimising a proxy signal. No ground truth label is needed in the training process. Most research either exploits geometrical cues such as multi-view geometry or epipolar geometry to learn depth One of such algorithm is unsupervised monocular depth estimation with Left-Right Consistency(monodepth) which is an monocular depth estimation algorithm [4]. This algorithm can perform even in absence of ground truth depth b hallucinating the depth for a given image and projecting it into nearby views, the model is then trained by minimizing the image reconstruction error. Another algorithm that determines a depth map from a single RGB input image is AdaBins Depth Estimation using Adaptive Bins [2] which uses a transformer based architecture block that divides the depth range into bins whose center value is estimated adaptively per image. The general idea that this paper uses is to perform a global statistical analysis of the output of a traditional encoder-decoder architecture and to refine the output with a learned post-processing building block that operates at the highest resolution. Using estimated depths and distance from the above defined models we can use simple physics based models to determine the power of each of the eye and verify it with the given prescription as close as possible.

4. Methodology

4.1. Dataset used

The 300-W is a face dataset that consists of 300 Indoor and 300 Outdoor in-the-wild images. It covers a large variation of identity, expression, illumination conditions, pose, occlusion and face size. The 300-W database contains a larger percentage of partially-occluded images and apart from covering normal expressions like smile, neutral, it also covers more expressions such as “surprise”,”terrified”,”depressed”,“scream”. The images were downloaded from google.com by making queries such as “ceremonies”, “conference”, “protests”, “cricket” and “celebrities”.Images were annotated with the 68-point mark-up using a semi-automatic methodology.

4.2. Procedure

A convolutional neural network (CNN) is a type of artificial neural network used in image recognition and processing that is specifically designed to process pixel data.Examples of CNN in computer vision are face recognition, image classification etc. It is similar to the basic neural network. CNN also have learning parameters like i.e, weights, biases etc in the neural network.The main technique in CNNs is convolution, where a filter slides over the input and merges the input value + the filter value on the feature map. Face landmark detection is a computer vision task where we want to detect and track key-points from a human face.A Convolutional Neural Network is trained on i-Bug 300-W facial Landmark dataset to get facial landmark coordinate points on face.Using this technique coordinates of points representing both eyes are calculated which is used to find cornea to cornea distance.Then depth is calculated using depth sensing model and calibration techniques.Using all these information and applying basic physics formula of eye power, the eye power is calculated.

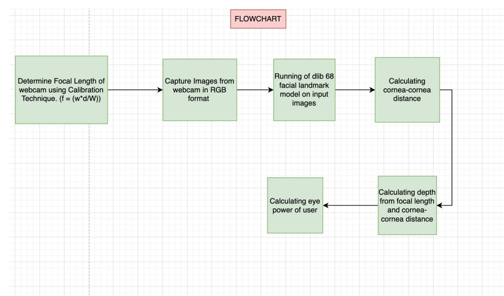

Figure 2. FlowChart

5. Experiments

Here we tested our model against two different individuals who had -2.5D and 0D power respectively.The results proved to be almost accurate with a small range of error which can be neglected.The images shown below display the same.

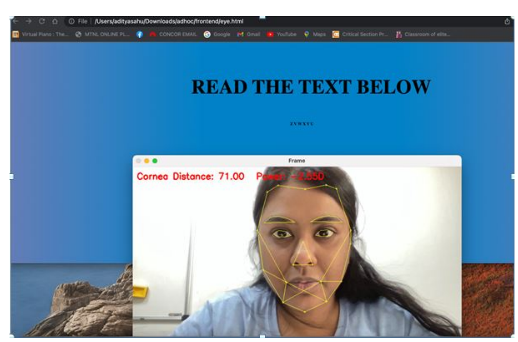

Figure 3. This test subject had -2.5D power

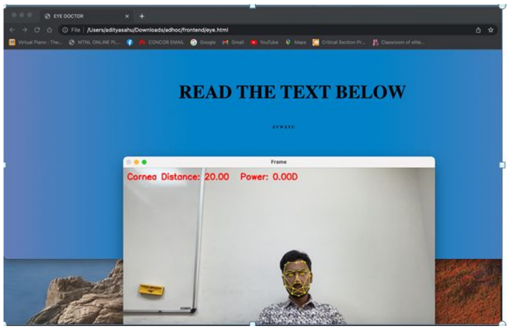

Figure 4. This test subject had 0D power.

6. Conclusion

In this project we are not proposing a model to eradicate the process of detecting/checking eye power which is currently prevalent in the healthcare centres.Our application can be used as a verification system over the existing vision testing procedures.Also the overall accuracy of our system can be further improved if a more accurate depth-sensing model can be developed for measuring the eye-camera distance.

7.

References

[1] Troy Bedinghaus. How to read an eyeglass prescription, May 2020. 1

[2] Shariq Farooq Bhat, Ibraheem Alhashim, and Peter Wonka. Depth estimation using adaptive bins. November 2020. 2

[3] R. Garg, G. Carneiro, and Reid. Unsupervised cnn for single view depth estimation: Geometry to the rescue, 2016. 2

[4] Clément Godard, Oisin Mac Aodha, Michael Firman, and Gabriel J. Brostow. Digging into self-supervised monocular depth prediction. October 2019. 2

[5] Darl Tan. Depth estimation: Basics and intuition. 2