Development of a Computer Vision eye gaze tracking algorithm for Virtual Reality applications.

Leopold Hebert-Stevens
CentraleSupelc
Université Paris-Saclay

Adrien Loizeau
CentraleSupelc
Université Paris-Saclay

1. Motivation and Problem Definition

Eye gaze estimation has become a very important feature implemented in many devices, such as assistive tools for impaired people, commercial contexts, or car manufacturing. A specific application of eye gaze tracking is in the Virtual and Mixed Reality (XR) field, where it enables energy-efficient rendering and gives the user a more realistic interaction with the content. By tracking the user's gaze, the VR headset can render only a specific part of the screen where the user is looking and thus improve the quality of that segment. [1]

However, multiple publications estimate that the results of direct end-to-end deep learning methods for gaze estimation are still not competitive with hybrid learning and geometric approaches. [4] Moreover, deep learning requires high processing computational power, which is very limited in current VR headsets. Therefore, this project aims to develop an efficient state-of-the-art eye gaze tracking algorithm from classic computer vision approaches and methods for Virtual Reality applications.

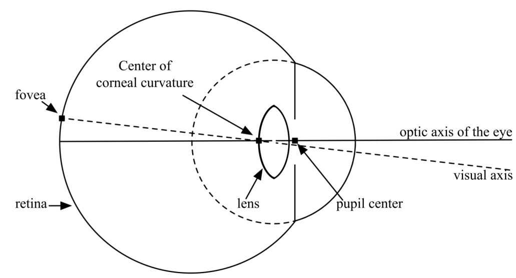

Figure 1. Schema of the geometries and structure of the human eye.

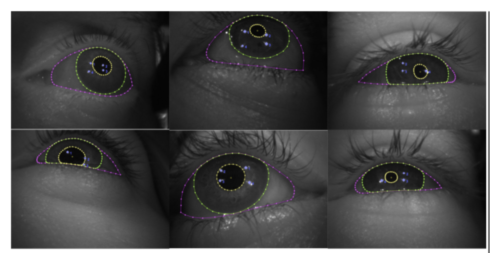

Figure 2. Examples of segmentation of the iris, pupil, and cornea of the images from the MagicEye dataset.

2. Methodology

For this project our plan is to apply both a geometrical and a structural approach to estimate the cornea center that gives us a gaze estimation.
Using a data-set of eye images, we would found different mathematical algorithms to apply some segmentation to the eye images and therefore locate important features such as the eye position. Image 2 shows a representation of such a segmentation. Being able to clearly estimate the position of the eye, the iris and the pupil in a two dimension plan would be a first real success.

The main steps of our project are first of all found the dataset we want to use to train and test our model. This dataset should be diverse enough to work represent any possible eye position and shape. Ideally, we would select a dataset with multi angles for each eye which would allow us to have a three dimension view of the eye and therefore estimate more precisely the gaze. A second step would be to treat the image and execute some computer vision algorithms to segment it. This step will be the most important of our project and will require us both intuition and readings.
A first idea for segmentation would be to do color detection to found the iris of the eye and then combine it with edge detection in order to found the iris and the pupil. Once this major step done a final one would be to combine many images of the same eye to get a 3D view of the eye. Such a view could be useful to better estimate the position of the cornea center of the eye.

We plan to mainly develop algorithms centered around eye segmentation. To do so we will use filtering and edge detection algorithms [3]. These ones will allow us to isolate the eye, the iris and the pupil. Such tools already exist but we still need to test them and modify them for our purpose. A second part of algorithms centered around the 3D estimation of the cornea center could be used, but we don't know now which ones.

According to the papers we read, today's state of the art models are based on CNNs but are still using a lot of traditional computer vision. In fact, end-to-end deep learning approaches aren't delivering good performances and the computer visions ones are still very competitive as they don't usually require too much computation power. We will therefore easily found prior work on this subject.

3. Evaluation

Our results will be evaluated on multiple criteria. First, we will consider the accuracy of the segmentation of the different parts of the eye, such as the cornea, the pupil, or the eyelids. Then we will compare the geometric computations of the cornea, pupil and gaze estimation to the ground truth of our dataset. We can also measure our CV model's reliability against different types of luminosity and camera quality by voluntarily degrading the dataset with filters or noise.

Many computer vision applications require eye focus; therefore, multiple eye-tracking datasets are available online. In the case of virtual reality, the cameras recording the eyes of the user are often located in the headset very close to the eye; and usually record in infrared frequency producing a black-and-white image. Therefore, we need to choose specific datasets that match our conditions. Many are available online but a very interesting and useful dataset is the one provided by MagicEyes MagicleapInc. [2] It is a very suited large-scale eye dataset collected using head-mounted VR devices, containing 880.000 images from 587 subjects. Moreover, 80,000 of the images feature manually annotated ground truth in 3D and segmentations. The rest of the images feature gaze target labels. Because this dataset presents high similarity to real applications, it will be valuable to verify the reliability of our model.

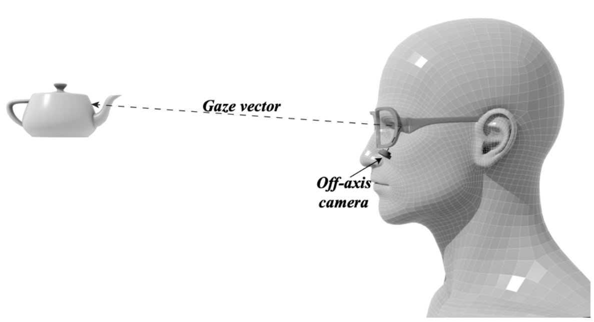

Figure 3. Representation of the position of the camera and the gaze vector of the user in a typical VR headset.

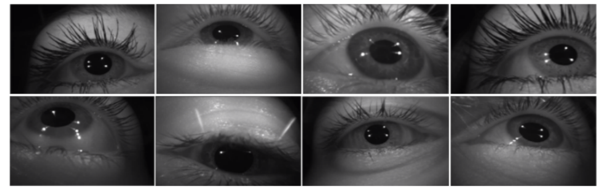

Figure 4. Examples of the images from the Magiceye dataset.

Characteristics	Types	Train Set	Test Set D1	Test Set D2
Gender	Male	51.2	48.4	46.7
	Female	48.8	51.6	53.3
Age	18-24	9.0	6.5	15.6
	25-34	36.2	31.2	30.2
	35-44	29.3	33.3	26.9
	45-54	19.5	20.4	24.0
	55-59	6.0	8.6	3.3
Ethnicity	Caucasion	58.4	54.8	59.0
	African Descent	23.7	29.0	17.0
	Hispanic	12.3	10.8	15.1
	Other	5.6	5.4	8.9
Eye Color	Brown	63.7	64.6	53.3
	Blue	10.2	15.1	19.8
	Hazel	17.1	14.0	19.8
	Green	9.0	4.3	7.1
Skin Color	Light	57.8	51.6	62.3
	Medium	15.0	17.2	15.1
	Brown	8.1	6.3	8.5
	Black	19.1	26.9	14.1
Mascara	Yes	21.0	17.2	22.2
	No	79.0	82.8	77.8

Figure 5. Demographic distribution from the MagicEye dataset.

Data Type	Train Set	Test Set Device 1	Test Set Device 2
Subjects	334	93	160
Images	62K	18K	800K
Segmentation	62K	18K	No
Pupil Center	62K	18K	No
Gaze Target	62K	18K	800K

Figure 6. Label availability for each ground truth type from MagicEye dataset.

References

[1] V. Clay, P. König, and S. Koenig. Eye tracking in virtual reality. *Journal of Eye Movement Research*, 12, 04 2019.

[2] Datagen. Eye datasets, 2021.

[3] A. Villanueva, J. J. Cerrolaza, and R. Cabeza. Geometry issues of gaze estimation. In S. Pinder, editor, *Advances in Human Computer Interaction*, chapter 30. IntechOpen, Rijeka, 2008.

[4] T. v. A. J. Z. V. B. Zhengyang Wu, Srivignesh Rajendran and A. Rabinovich. Magiceyes: A large scale eye gaze estimation dataset for mixed reality, 2020. Magic Leap Inc, USA.