

Operator's Gaze Direction Recognition

Alexey Popov¹, Vlad Shakhuro¹ and Anton Konushin^{1,2,3}

¹Lomonosov Moscow State University, 1, Leninskie Gory, 119991, Moscow, Russia

²National Research University Higher School of Economics, 11, Pokrovsky boulevard, 109028, Moscow, Russia

³Samsung AI Center, 10, Butyrskiy Val Ulitsa, 127055, Moscow, Russia

Abstract

This work is devoted to the algorithm for recognizing the direction of a operator's gaze. The paper considers the method of classifying the direction of gaze. The algorithm proposed in this paper allows to determine the driver's gaze direction, which helps to protect the driving process. In the proposed work, we reviewed the algorithms and methods used to recognize the human gaze in various conditions. For each given method, a list of their positive and negative qualities is indicated. Based on the results of the review, an algorithm for classifying the car driver's gaze direction is proposed. The proposed algorithm consists of two components. The first part is responsible for the regression of the gaze direction. The second part is responsible for classifying the results of the first part. Experimental evaluation of the developed algorithm has shown that it is effective for the task of classifying the gaze direction. An important advantage of this algorithm is that it does not require repeated training of the algorithm to adapt to other scenarios of gaze direction classification.

Keywords

Gaze direction recognition, Convolution neural network, Driver gaze classification, Gaze recognition

1. Introduction

The task of the operator's gaze direction recognizing is key in many computer vision systems, such as driver assistance systems, analyzing human attention, determining the most interesting contextual data (for example, on a *WEB*-page), intelligent compression of media information, and many others. Thus, the creation of a new, more accurate and powerful algorithm for the gaze direction recognizing will improve the quality of these systems.

In this paper we propose an algorithm for solving the problem of classifying the car's driver gaze direction. Nowadays systems using computer vision algorithms are often used to ensure active security, for example: *ADAS*. The system for classification of the operator's gaze, in particular the driver of the vehicle, is necessary for ensuring active safety, as it can prevent the driver from being dangerously distracted from the process of driving the vehicle.

In many countries, there are a number of rules that prohibit drivers of motor vehicles from using the phone while driving, to prevent mass violations of such rules, driver's gaze detection


GraphiCon 2021: 31st International Conference on Computer Graphics and Vision, September 27–30, 2021, Nizhny Novgorod, Russia

✉ alexey.popov@graphics.cs.msu.ru (A. Popov); vlad.shakhuro@graphics.cs.msu.ru (V. Shakhuro); anton.konushin@graphics.cs.msu.ru (A. Konushin)

🆔 0000-0001-7569-1369 (A. Popov); 0000-0002-1586-9257 (V. Shakhuro); 0000-0002-6152-0021 (A. Konushin)



© 2021 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

system also can be used, determining whether the driver is looking at the road or distracted. If the driver is distracted by the smartphone, warn him about the violation or fine him.

Also, the algorithm for classifying the direction of the driver's gaze can be used to build a rating of taxi drivers and carsharing, which will ensure the safety of taxi passengers and businesses in the field of carsharing.

To solve the problem of classifying the gaze direction, the use of automatic tools is required. Computer vision algorithms are actively developing, aimed at solving the problem of classifying the gaze direction without using additional technical means. The introduction of such algorithms into the daily life of drivers will increase the level of driving safety.

At the moment, most of the methods used to determine the gaze direction, namely its classification or regression of a three-dimensional angle, use classical approaches and do not resort to the use of powerful, modern neural network algorithms. Based on this, the development of a neural network algorithm for recognizing the direction of the operator's gaze is a promising task, since neural network methods are much more powerful and accurate than classical methods.

There are single-stage neural network algorithms for classifying the gaze direction, which allow obtaining high-quality results for a fixed work scenario. To apply such algorithms in new work scenarios, a complete retraining of the algorithm is required, which is a big problem. This problem is especially noticeable in the problem of classifying the direction of the driver's gaze, since car drivers can often change, the driver's position in the seat can change, and the target set of classes can change depending on the scenario of using the algorithm. Therefore, the task of developing a two-stage algorithm that allows adapt to new scenarios of using the method without retraining.

In the paper we propose a method for classifying the direction of the operator's gaze, according to the data from the camera that captures the operator, namely the driver of a motor vehicle. The proposed method for determining the gaze direction consists of several important parts. To localize the operator's face, a face detector is used, and a neural network with a new architecture for this task is used to determine the vector of the three-dimensional gaze angle. Using the results of the algorithm for determining the direction of the three-dimensional gaze angle, it is possible to classify the gaze directions on the zones, which solves the chain that is used to solve the problem of classifying the gaze direction in this work. The proposed method was tested on several different data sets, including a self-assembled data set. As a result, the implemented algorithm of driver's gaze direction classification allows to achieve high accuracy in different scenarios of driver's gaze direction classification.

2. Related work

The existing methods for determining the direction of the gaze are divided into two categories, the first of which is the methods for classifying the direction of the gaze, and the second is the methods for regressing the direction of the gaze.

Consider the existing methods for recognizing the gaze direction:

- Methods that classify the gaze direction to predefined sectors in a specific work scenario.
- Methods that determine the three-dimensional gaze angle.

2.1. Gaze direction classification algorithm

The task of classifying the gaze direction arose quite a long time ago, so there are a lot of different approaches to solving this problem, in each work the authors offer their own heuristics and their own experimental conclusion.

To solve the problem of classifying the gaze direction, eye models are often developed, since the eye is simply described as a geometric object. In the article [1], the authors suggest using classical methods and a two-stage algorithm. In the first stage of the algorithm, the area containing the face is extracted from the image using *Viola Jones*, and in the second stage using *CLNF (Constrained Local Neural Field)* key points inside the eye are extracted. Using the data obtained at the second stage of the algorithm, it is necessary to obtain feature vectors that can be used to solve the classification problem.

In the article [2], the authors propose a neural network method for classifying the gaze direction, they take the *Dlib-ml* detector of the face and key points of the face, and then determine the position of the eyes. After that, both eyes are fed to the neural network and it builds the probability distribution of the classes of the gaze direction from the two images.

One of the first articles in which the authors proposed to replace the main stage of the algorithm with a neural network was the article [3]. In particular, in this paper, the algorithm for classifying the gaze direction was replaced by a neural network one. Since, according to the authors, the features extracted by the neural network from the image of the eye area can well classify the gaze direction, due to the greater power of neural algorithms.

There are approaches in which the authors try to implement the most qualitative method of gaze classifying, only by the image of the eye area. In such works, the authors pay more attention to the classification method, without being distracted by the task of determining the position of the eye and face. In the article [4], the authors prepared a data set containing placed images of the eye areas in advance and trained their own lightweight neural network classifier on them.

2.2. Three-dimensional gaze angle recognition algorithm

Developing an algorithm for recognizing a three-dimensional angle that determines the exact gaze direction is a difficult task. The accuracy of the existing methods increases every year. Now neural network algorithms are increasingly used for angle regression, since this approach can significantly increase the quality of the final method.

More and more complex neural network models are used to accurately determine the gaze angle. In the article [5], the authors propose to consider and submit to the network input 3 images at once, this is the image of each eye and face as a whole. The main idea of the authors is that the weight of each of these three images varies from case to case. In this way, the authors try to force the network to independently distribute attention between the input images. To do this, a weight regressor is added to the algorithm, which calculates the weight of each of the input images in the final result, namely, the calculated angle of gaze direction. In order to correctly assess the contribution of each of the images to the final result, the article proposes a new loss function that distributes the weight between the features of each of the input images in the final value of the viewing angle.

The old methods often use not only an ordinary camera, but also an infrared camera to reduce the influence of external factors on the operation of the algorithm, since the power of classical methods is much less than that of neural network methods. For example, the article [6] uses an infrared flashlight and a camera with a removed filter to make the lighting in the car more uniform and reduce the influence of external light sources, such as lights on poles, headlights of an oncoming car, sun glare.

Often, in solving the problem of recognizing the gaze direction, the main problem for authors is the fight against occlusions. For example, in a situation where the car is moving, the position and tilt of the head can change significantly, and the lighting can also vary greatly, for example on a sunny day when entering a tunnel. There are many algorithms in which the authors try to come up with a method to deal with the described problem. In the article [7], a new network architecture is proposed, in which an additional network is added before the regressor, designed to create a gaze direction map that is invariant in various scenarios of the algorithm application. To train this network, the authors use marked-up images that already have a gaze map, which imposes significant restrictions on the application of this method in real conditions.

2.3. Datasets used in gaze direction recognition tasks

To solve the problem of determining the gaze direction, a lot of data is required. This is due to the fact that the problem is mainly solved using machine or deep learning methods. Many data sets exist to solve this problem, but most of them are designed to solve highly specialized problems of classifying the gaze direction.

One of the main datasets with eye images is the *MPHIGaze* dataset proposed in the article [8]. It contains 213659 images depicting 15 different people.

Special methods for generating synthetic data sets are also often developed. One of the most popular synthetic datasets for training models for determining the gaze direction is [9].

There are various specific conditions in which the method of determining the gaze direction has to work, so there are many different forms of data sets. In general, data sets differ in the way they are collected and marked up, but there are also more significant differences. Some data sets are captured on an infrared camera, some are collected using a camera with depth data, such as the data set [10].

There are also simpler datasets that contain only color images of people's faces taken with a conventional camera. Such datasets are often more popular due to their simplicity and versatility. It is important to note that such data sets are easier to collect, which affects their quality and size. An example of such a dataset is *XGaze*.

Some data sets are collected using new markup tools, such as using the voice commands of a person who is in the field of view of the camera and is filmed to collect data. An example of such a data set is *Driver Gaze in the Wild*, it was collected using the method described in the article [11]. The *Driver Gaze in the Wild* contains 29050 images of 383 different people.

3. Gaze direction recognition

Based on the conclusions drawn from the review of methods and the task of constructing a universal method for determining the direction of the operator's gaze, the following scheme of

the method was developed.

The method proposed in this paper consists of two parts, the first of which is the regression of the gaze angle, the second is the classification of the obtained gaze angle for a certain scenario of the method.

3.1. Gaze direction regression

All methods for determining the direction of the gaze, in which there is a stage of predicting the vector of the direction of the gaze, are combined by a common task of choosing the target predicted vector and the metric in which the error of the method will be calculated during training.

3.1.1. Metrics and loss function

There are many different metrics (loss functions) for training neural networks, but each task has its own specifics, which significantly limits the choice. Later in this section, the metric used to evaluate the model is also a loss function when training this model.

In our problem, we can predict the gaze direction vector as a three-dimensional vector, namely $V = (X, Y, Z)$, where X, Y, Z are the coordinates of the gaze direction vector in three-dimensional space. In this problem, we only care about the directions of this vector, which allows us to use the cosine similarity metric (*cosine similarity*), which is expressed by the formula 1, where A and B are two vectors compared by this metric, and n is the number of elements in each of the vectors A and B .

$$\text{CosineSimilarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \cdot \|B\|} = \frac{\sum_{i=1}^n (A_i \cdot B_i)}{\sqrt{\sum_{i=1}^n A_i^2} \cdot \sqrt{\sum_{i=1}^n B_i^2}} \quad (1)$$

The main advantage of such a metric is that after training the method, there is no need to map the vector to a different dimension, and it is enough only to normalize the values that the neural network outputs. But in this context, we can also consider the disadvantages of this approach, namely: an increase in the number of network parameters and a lower speed of calculating such a metric on *GPU*. It is also worth noting that training a neural network with such a metric does not limit the length of the vectors that the network outputs, since the metric penalizes the method only for divergence in the direction of the vector, which in turn can cause large computational inaccuracies.

On the other hand, to solve the problem of determining the gaze direction, a spherical coordinate system with a fixed radius of the sphere can be used. In this case, it is necessary to predict the vector $G = (\phi, \theta)$, where ϕ, θ are angles in a spherical coordinate system, the notation of angles in a three-dimensional coordinate system can be found in the figure ?? . Using the angles ϕ and θ , as well as the system of formulas 2 with subsequent normalization of the vector $V = (X, Y, Z)$ by the formula $V = \frac{V}{\|V\|}$, it is possible to uniquely express the three-dimensional vector of the gaze direction.

$$\begin{cases} X = 1, \\ Y = X \cdot \tan(\phi), \\ Z = \sqrt{X^2 + Y^2} \cdot \tan(\theta), \\ V = (X, Y, Z) \end{cases} \quad (2)$$

The unambiguous of converting angles into a vector is confirmed by the fact that the vector of the direction of gaze is located in a positive half-space along the Ox axis.

Using the representation described above for the vector of the gaze direction has several advantages, including: reducing the number of network parameters, simplifying the loss function used, as well as optimizing the values of angles in some neighborhood of the true value, which does not allow the optimized response vector to increase indefinitely without affecting the value of the quality metric. In this approach, the quality metric MSE is best suited, which is described by the formula 3, where A and B are two vectors compared by this metric, and n is the number of elements in each of the vectors A and B .

$$MSE = \frac{1}{n} \cdot \sum_{i=1}^n (A_i - B_i)^2 \quad (3)$$

For a more objective comparison of the models and metrics described above, a basic neural network for predicting the vector of the gaze direction was developed to solve the problem of determining the three-dimensional angle of the gaze direction. With the help of this neural network, it was possible to compare the quality of these models.

It is also necessary to compare the quality of the base model on color and black-and-white images in order to optimize the hardware needs of the method described in this work. It can be argued that for the task at hand, black and white images contain all the necessary information.

Thus, at this stage of the method development, the following conclusions were made: the use of the method of predicting the angles ϕ and θ using the MSE metric shows the best quality, the use of black-and-white images for training the method allows to increase the final quality of the method.

3.1.2. Proposed method

This section describes the main idea of the method of regression of the gaze direction. To implement this idea, a new neural network architecture is proposed, which allows solving the task with high accuracy.

To determine the direction of the gaze, the proposed method uses an image of a person's face.

The reason for using the face image as a whole was the problem of using only eye images in real conditions of using the method. When using eye images, overlaps often occur and an error introduced by eye detection algorithms has a great influence. In the figure 1, there is an example of overlapping the eye with a pair of glasses in a real scenario of using the method of determining the direction of the gaze, such an example of an image is practically insoluble for the method of determining the direction of the gaze only by the eyes.

The method proposed in this paper has a simple idea-to solve the problem of determining the gaze direction without significantly increasing the requirements for input data. For this purpose,



Figure 1: Example of a complex image for the method of determining the direction of gaze by the eyes

an extensive neural network architecture was developed, the input of which is supplied only with a black-and-white image of the operator's face. This network architecture allows decrease of the error and the operating time of the method due to inaccuracies in eye detection, and also allows train the model using a high-quality data set *XGaze*.

The main purpose of the proposed architecture shown in the figure 2, is a one-stage fully neural network solution to the problem of determining the three-dimensional gaze angle. To do this, several architectures were tried, which include several different-scale paths within the network, which allow the network to focus on different-scale features. The architecture proposed in this paper uses three branches, two of which coincide in architecture and in the scale of the features under consideration, and the third is different and is designed to work with larger-scale features. Such an architecture, according to the assumption, should use two identical branches to isolate small-scale features for each of the eyes from the image, and the third branch should allocate larger-scale features that are responsible for the position of the face, which can be used to determine the direction of head rotation.

This architecture allows to get rid of a more complex scheme of the algorithm, when it is necessary to cut out the parts of the image that depict the eyes, and then submit three images at once to the network input, although with this approach, the complexity of the neural network architecture will not decrease.

In this paper, an alternative approach is also tested, a neural network that receives three images at once: a face, a left eye, and a right eye. This approach does not win over the basic approach described above in terms of the complexity of the neural network architecture, since it also contains three branches. To implement the approach with multiple input images, firstly, it is necessary to select parts of the face image that contain the eyes. To solve the problem of eye detection, it is necessary to involve a separate method that will supply additional noise to the described algorithm. Based on the facts described above, we can conclude that the alternative

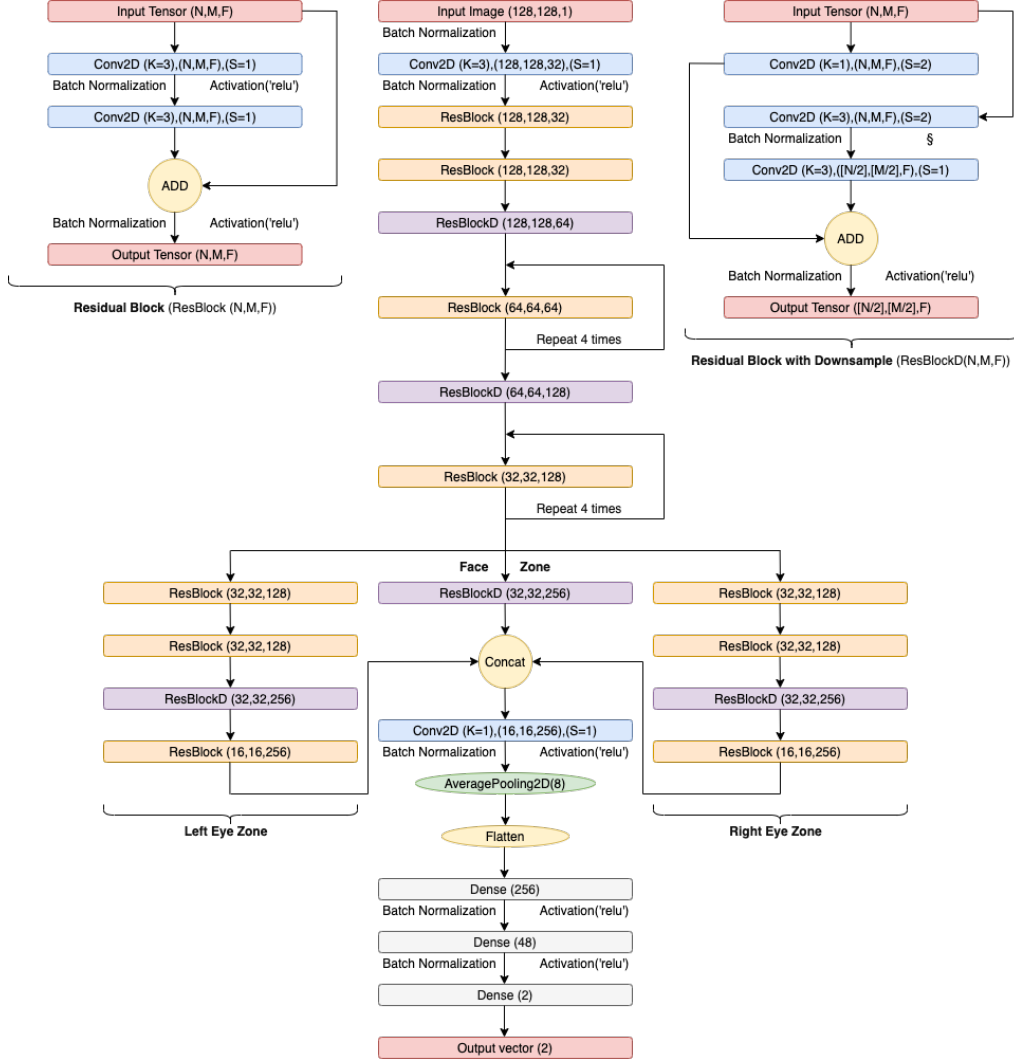


Figure 2: The main architecture for determining the gaze direction from the face image (K is the size of the convolution core, S is the convolution step)

approach does not win in terms of complexity and speed, and also requires additional image processing methods.

3.2. Gaze direction classification

Since the proposed method pursues the goal of achieving a high level of universality, namely, the possibility of applying the method in various scenarios without retraining. To classify the gaze direction, a three-dimensional angle of the gaze direction is used, determined at the first step of the algorithm.

It is necessary to solve the problem of classifying pairs of the form (ϕ, θ) for fixed classes, since from the pair (ϕ, θ) the three-dimensional gaze angle is uniquely determined. The number

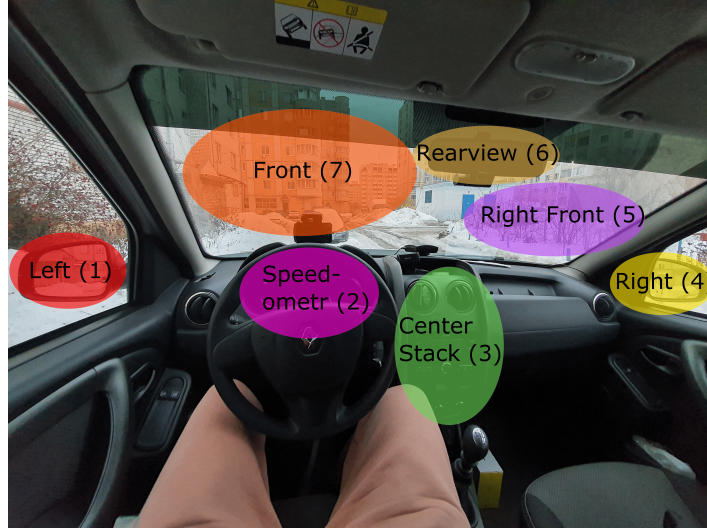


Figure 3: Class locations proposed in this paper and in the dataset *DriveDS*

of zones into which the field of gaze direction can be divided can be arbitrary, but in this work we will consider the classification of the driver's field of gaze into 7 zones: left mirror, speedometer, center console, right mirror, right part of the windshield, interior rearview mirror, left part of the windshield. An example of the location of classes in the figure 3.

To solve the classification problem, several different classical methods are tested in this article. For this classification problem, the neural network approach is not very well applicable, since the geometry of the classes is quite simple, and the amount of training data is small, so the classical methods were also tested: *Nearest Centroid*, *K Nearest Neighbors*, *Tree Classifier*.

The *K Nearest Neighbors* method was chosen as the basic one, according to the set of positive qualities for the problem being solved, since this method allows for a small amount of training data with a fairly simple arrangement of classes to obtain a higher classification quality compared to other methods.

To apply the proposed method, it is proposed to calibrate the classification method for each new car driver, in order to look at each fixed zone for 2 seconds.

4. Experimental evaluation

All neural networks were trained using an Nvidia Geforce GTX 1080 Ti graphics card. The basic model of the regression of the gaze direction was trained on the data set *UnityEyes* from the article [9] using the same optimizers. The training of the classification method was carried out on the Drivers data set, which was prepared as part of the work on this article.

Table 1Comparison of 3D angle errors of different architectures on the dataset *XGaze*

Method	3D angle error
Proposed architecture	3.14
Architecture from paper [5]	3.41
Architecture from paper [12]	4.50

Table 2Comparison of the accuracy of different classifiers on a part of the data set *DriveDS*

Method	Accuracy
<i>Nearest Centroid</i>	90.0%
<i>K Nearest Neighbors</i>	93.1%
<i>Tree Classifier</i>	88.2%

4.1. Gaze direction regression

In this paper, several architectures are proposed for determining the three-dimensional angle of the gaze direction from the image to the face. The main architecture is shown in the figure 2.

The comparison of the quality of the architectures proposed in this paper can be found in the table 1, also in this table the results obtained by the authors of the article [12] on the dataset *XGaze* are presented, on which the quality of the models is compared.

According to the results in the table 1, it can be concluded that the proposed basic architecture, shown in the figure 2, shows the best results at the moment, while surpassing the result obtained by the authors of the dataset *XGaze*.

4.2. Gaze direction classification

In the proposed work, several different classifiers of the gaze angle are compared. In the 2 table, there are the results of each of the classifiers on the part of the *DriveDS* data set that shows one person.

The results in the table 2 are obtained on classifiers that are trained on 120 random examples of angles from each class. It also follows from the presented table that the *K Nearest Neighbors* method is best suited for solving this problem, which shows the highest results. An important advantage of *K Nearest Neighbors* is the high speed of operation.

For a more detailed description of the classification methods, we present the class distribution schemes on a two-dimensional plane, Figure 4.

Also, for the classification method *K Nearest Neighbors*, we give an error matrix for a part of the data set *DriveDS* that shows one person, the table 3.

The work carried out a cross-comparison of the method's work on different people from the data set *DriveDS*, the classification method was trained on the examples of one of the three people, and then tested on the full data set of each of the people. The results of cross-testing are shown in the 4 table.

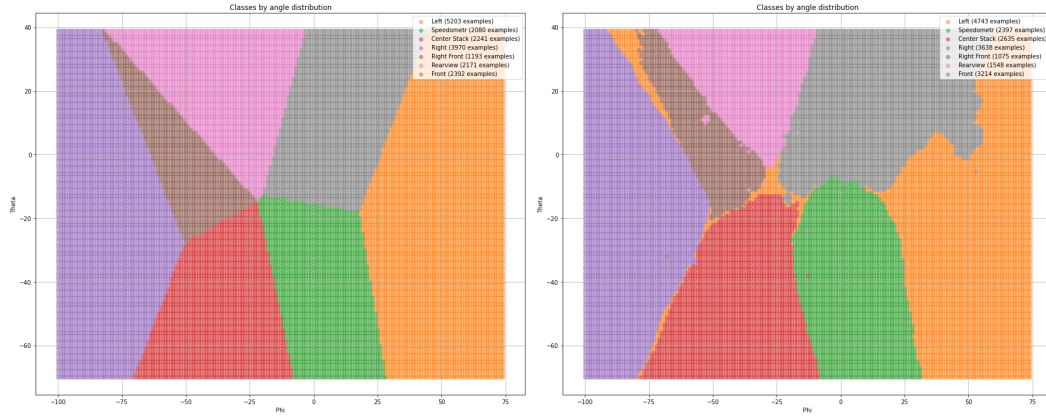


Figure 4: Division of a two-dimensional plane into classes by various methods of classifying the gaze angle (**Left** - Nearest Centroid, **Right** - K Nearest Neighbors)

Table 3

Confusion matrix for the *K Nearest Neighbors* method on the part of the data set *DriveDS* (**Left**-Left, **Speed**-Speedometer, **CntCons**-Center console, **Right**-Right, **R Front**-Right Front, **Rear**-Rearview, **Front**-Front, example in figure 3)

True \ Predict	Predict						
	Left	Speed	CntCons	Right	R Front	Rear	Front
Left	97.3	0.0	0.0	0.0	0.0	0.0	2.4
Speed	0.0	96.4	0.0	0.0	0.0	0.0	2.8
CntCons	0.0	1.6	95.6	0.0	1.3	0.0	0.0
Right	0.0	0.0	1.2	98.0	0.0	0.0	0.0
R Front	1.1	0.0	1.5	3.8	89.9	2.9	1.0
Rear	1.1	0.0	0.0	0.0	4.9	90.8	2.8
Front	6.1	5.7	0.0	0.0	0.0	2.2	85.3

Table 4

Results of cross-testing on different people from the dataset *DriveDS*

Train \ Test	Test		
	Person1	Person2	Person3
Person1	93.2%	80.1%	84.6%
Person2	85.5%	85.3%	75.3%
Person3	89.5%	74.0%	89.3%

The results in the table 4 indicate that the method proposed in this paper allows it to be applied to various people without significant quality losses. It can also be concluded from the table that the calibration of the method does not always give a significant increase in quality, which means that universal parameters can be used for several drivers of the vehicle.

The average speed of the method is 6 frames per second on the graphics card *Nvidia Geforce GTX 1080 Ti*.

5. Conclusion

The algorithm proposed in this article allows us to obtain high accuracy of the classification of the operator's gaze direction using a two-stage algorithm, the first stage of which is the regression of the gaze direction, and the second is the classification of the obtained gaze direction vector. The main advantage of the proposed method is a high level of versatility. There is no need to retrain the method to switch to a new scenario of the classification algorithm.

References

- [1] N. H. Jabber, I. A. Hashim, Robust eye features extraction based on eye angles for efficient gaze classification system, in: 2018 Third Scientific Conference of Electrical Engineering (SCEE), 2018, pp. 13–18. doi:10.1109/SCEE.2018.8684107.
- [2] D. Saha, M. Ferdoushi, M. T. Emrose, S. Das, S. M. M. Hasan, A. I. Khan, C. Shahnaz, Deep learning-based eye gaze controlled robotic car, in: 2018 IEEE Region 10 Humanitarian Technology Conference (R10-HTC), 2018, pp. 1–6. doi:10.1109/R10-HTC.2018.8629836.
- [3] A. George, A. Routray, Real-time eye gaze direction classification using convolutional neural network, in: 2016 International Conference on Signal Processing and Communications (SPCOM), 2016, pp. 1–5. doi:10.1109/SPCOM.2016.7746701.
- [4] X. Wu, J. Li, Q. Wu, J. Sun, Appearance-based gaze block estimation via cnn classification, in: 2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP), 2017, pp. 1–5. doi:10.1109/MMSP.2017.8122270.
- [5] X. Zhou, J. Jiang, Q. Liu, J. Fang, S. Chen, H. Cai, Learning a 3d gaze estimator with adaptive weighted strategy, IEEE Access 8 (2020) 82142–82152. doi:10.1109/ACCESS.2020.2990685.
- [6] F. Vicente, Z. Huang, X. Xiong, F. De la Torre, W. Zhang, D. Levi, Driver gaze tracking and eyes off the road detection system, IEEE Transactions on Intelligent Transportation Systems 16 (2015) 2014–2027. doi:10.1109/TITS.2015.2396031.
- [7] S. Park, A. Spurr, O. Hilliges, Deep pictorial gaze estimation, Lecture Notes in Computer Science (2018) 741–757. URL: http://dx.doi.org/10.1007/978-3-030-01261-8_44. doi:10.1007/978-3-030-01261-8_44.
- [8] X. Zhang, Y. Sugano, M. Fritz, A. Bulling, Appearance-based gaze estimation in the wild, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 4511–4520. doi:10.1109/CVPR.2015.7299081.
- [9] E. Wood, T. Baltrusaitis, L.-P. Morency, P. Robinson, A. Bulling, Learning an appearance-based gaze estimator from one million synthesised images, 2016, pp. 131–138. doi:10.1145/2857491.2857492.
- [10] R. F. Ribeiro, P. D. P. Costa, Driver gaze zone dataset with depth data, in: 2019 14th IEEE International Conference on Automatic Face Gesture Recognition (FG 2019), 2019, pp. 1–5. doi:10.1109/FG.2019.8756592.
- [11] S. Ghosh, A. Dhall, G. Sharma, S. Gupta, N. Sebe, Speak2label: Using domain knowledge for creating a large scale driver gaze zone estimation dataset, 2021. arXiv:2004.05973.
- [12] X. Zhang, S. Park, T. Beeler, D. Bradley, S. Tang, O. Hilliges, Eth-xgaze: A large

scale dataset for gaze estimation under extreme head pose and gaze variation, 2020.
`arXiv:2007.15837`.