

EYE RECOGNITION SYSTEM TO PREVENT ACCIDENTS ON THE ROAD

N.A. Andriyanov^{1*}, A.A. Lutfullina¹

¹Financial University under the Government of the Russian Federation, 125993 Leningradskiy pr-t, h. 49, Moscow, Russia -
nikita-and-nov@mail.ru

Commission II

KEY WORDS: Face Detection, Face Recognition, Eye Detection, Eye Recognition, Convolutional Neural Networks, Haar Detector, Recall, Skip Target

ABSTRACT:

Today, possibilities of artificial intelligence allow us to see the emergence of autonomous cars. However, there are still many problems in this area at present. Often, such vehicles are "too slow to think", are not able to reliably process data from video cameras in the event of reflections, glare, and there are also questions about the safety of such driving in difficult weather conditions or in heavy traffic. At the same time, the human factor plays a major role in accidents of driven vehicles. Many accidents involve driver fatigue, distraction, or even falling asleep. At the same time, it is potentially possible to monitor the state of a person behind the wheel by a video sequence received from a camera installed in the car's interior and registering the driver's face in video sequence. In this paper, the existing databases of images of faces and eyes are considered, and an algorithm is presented that detects the state of closed eyes based on Haar detectors and convolutional neural networks.

1. INTRODUCTION

Now days one of the main applications for artificial intelligence, along with natural language processing and reinforcement learning, is computer vision. Indeed, a person receives about 70% of information through the organs of vision, so the appearance of many applications with technical vision is obvious.

At the same time, one of the areas of application of computer vision algorithms is the detection, recognition and analysis of a person directly.

Most often, this task comes down to detecting a face in the image and comparing this face with faces from the existing database. However, in some tasks, recognition of a specific person may not be required. In particular, the faces of different people have many of the same features. In this case, a person's face can be used to detect parts of the face, such as nose, mouth, eyes, as well as to determine emotions expressed on the face.

The task of highlighting the eyes, in turn, is relevant not only in various monitoring systems (distraction during classes, averting eyes during an online exam, etc.), but also in the analysis of driver states. According to statistics, the majority of road traffic accidents are related to the human factor, and about 20% of all cases are associated with falling asleep while driving (Kamberova, 2017).

At the same time, modern algorithms for pattern recognition (Andriyanov, 2020; Wang, 2021; Choy 2020) allow, based on the use of convolutional neural networks (CNNs) and deep learning (DL), to obtain sufficiently convincing characteristics in terms of recognition accuracy. Together with the methods of object detection (Andriyanov, 2019; Beal, 2020; Yu, 2021), it seems possible to construct an effective algorithm that will recognize the states of fatigue and sleep by the video sequence of the driver's face. Based on the results of such recognition, it

is possible to carry out some kind of counter-effect on the driver, for example, by generating a loud sound signal.

2. RELATED WORKS

Fatigue and falling asleep of drivers is the main cause of road accidents. This topic is really so relevant that car companies, together with scientists, are creating various systems that can analyze driving habits, and some even analyze brain waves and vital functions of the driver. Most of these algorithms are developed using machine learning. All of them can be classified into 3 categories.

Algorithms of the first category set the main goal to find changes in the behavior of a vehicle. An example of such an algorithm is described in (McDonald, 2018). The authors obtained a model that reveals anomalies in steering wheel deflection, vehicle speed and gas pedal position. The analysis of such anomalies is performed using a Bayesian network, the output of which decides whether the driver is sleepy or tired. The algorithm showed a low false alarm probability compared to methods (Johnson, 1998) predicting fatigue by assessing eyelid movements. It should be noted that the algorithm makes the decision as a result of the strong context of the situation, due to which false positives were reduced.

The second category of algorithms processes the vital signs of drivers, brain waves and electroencephalogram (EEG) readings. Based on the analysis of such time series, a forecast of the future state is made. Work (Wei, 2018) is devoted to the comparison of various EEG devices. The main result of the study was the conclusion about the possibility of using new convenient devices. However, the EEG itself does not provide an accurate identification of the stages of drowsiness. Work (Kartsch, 2018) supplements the studies of EEG signals with sensors of inertial devices. As a result of such aggregation, very high metrics are obtained, up to 95% accurate sleep detection. But the authors of the work (Tateno, 2018) proposed to monitor the heart rate and respiration of a person. However, the use of

additional devices while driving is not currently seen as a promising direction.

Finally, the last group of methods is computer vision-based methods. It is important to note that the driver's facial features change a lot when he gets tired, which allowed the development of a convolutional neural network for analyzing the driver's condition. Closure of the eyes was proposed to determine by measuring the angle of curvature of the eyelid (Tayab, 2019). The achieved accuracy is up to 95%, but high-quality shooting conditions and high image resolution are required. In (Shakeel, 2019), the application of the MobileNet-SSD architecture with training on 350 images (custom approach) was investigated. The main advantage was the ability to use the system on Android, on the other hand, the accuracy was about 80%. Additional studies (Celona, 2018) significantly complicate the processing device, since they analyze the eyes, mouth, and head posture.

Thus, in our opinion, an interesting approach is using computer vision, which requires an increase in metrics and an increase in productivity.

3. METHODS AND DATASETS FOR FACE AND EYE DETECTION AND RECOGNITION

When creating a face recognition system, among the set of algorithms, one can single out the general structure of the process (Kolomiets, 2014), shown in Figure 1:

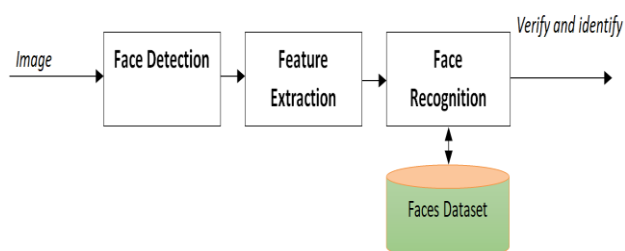


Figure 1. The structure of the face recognition system

At the first stage, preprocessing, detection and determination of the location of a human face in the image takes place, for this, methods such as the Viola-Jones detector, the histogram of oriented gradient (HOG) and the principal component analysis (PCA) are used.

At the second stage, a vector of features is allocated for each face previously detected. HOG, Eigenface, independent component analysis (ICA), linear discriminant analysis (LDA), scale invariant function transform (SIFT), Gabor filter (GF), local phase quantization (LPQ), Haar wavelets (HW), Fourier transforms (FT), and local binary pattern (LBP) widely used to highlight facial features (Kolomiets, 2014).

At the final stage, the found feature vectors are compared with the available features in the database to find the most likely similarity between them for further identification of a person, it is known that correlation filters (CF), convolutional neural network (CNN), and the k-nearest neighbors method (K-NN) effectively solve this problem.

Compared to other biometric human identification systems, such as eye, iris, or fingerprint recognition, facial recognition is

not the most efficient or reliable. Moreover, when identifying a face in a natural environment, there are many problems associated with different lighting, facial expressions, age or dynamic background. Therefore, the review will focus on more advanced methods of face recognition using various databases. All considered methods can be classified according to the used features of objects into three groups (Kuznetsov, 2020):

1. Structural approach - describes objects as a system consisting of many interconnected elements (eyes, nose, mouth, etc.), includes methods such as scale-invariant feature transform (SIFT), SURF (Speeded Up Robust Features), BRIEF (binary robust independent elementary features), local binary pattern (LBP), histogram of oriented gradients (HOG), local phase quantization (LPQ).
2. Holistic approach describes objects as a whole, projecting the entire image into a smaller subspace or correlation plane, includes such linear methods as principal component analysis (PCA), linear discriminant analysis (LDA), Eigenfaces, and nonlinear Kernel PCA (KPCA), convolutional neural network (CNN), support vector machine (SVM).
3. Hybrid (Cluster) approach combines structural and holistic approaches to improve recognition accuracy.

Among the hybrid methods, it should highlight the hybrid CNN-LSTM-ELM system, invented by a group of Chinese scientists (Sun, 2018), which perfectly recognizes human activity (HAR). Figure 2 shows a visual diagram of this process.

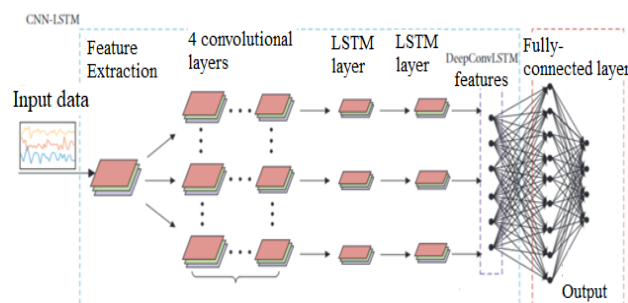


Figure 2. The structure of the face recognition system



Figure 3. Datasets containing images of faces



Figure 4. Datasets containing images of eyes

Further Figure 3 and Figure 4 show datasets available in the public domain and used when working with faces and / or eyes in images.

Thus, it was decided to use the Viola-Jones algorithms and Haar detectors to extract faces in images from face databases, followed by geometric selection of eyes in the face area, as well as images with eyes directly.

4. TRAINING A NEURAL NETWORK USING READY-MADE DATASETS

Since face recognition task has been modified into the problem of recognizing the state of the eyes, a logical change in the scheme shown in Figure 1, became the diagram shown in Figure 5.

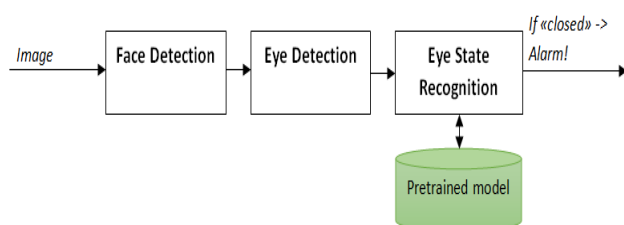


Figure 5. Scheme of the eye condition analysis

It can be seen from the presented diagram that it works for one eye. In fact, 2 pre-trained models have been allocated, one for the left eye and the other for the right eye. CNNs for such models were trained separately. From the available database for training and image testing, a sample was created containing 10,000 images, of which 2,000 are a closed right eye for training, 2,000 are a closed left eye for training, 2,000 were an open right eye for training, 2,000 are open left eye for training, and 500 images of each of the four states described earlier for testing. The bases considered earlier were used, and the resulting images of the eyes were converted to 28x28 pixels. In both cases, training took place on the basis of its own neural network with 5 layers of 128 neurons each and with the training transfer of the VGG-19 network (Koonce, 2021). Figure 6 shows the architecture of the VGG-19 neural network.

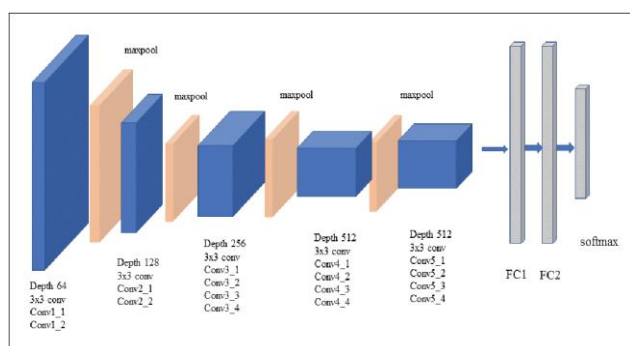


Figure 6. VGG-19 architecture

Figure 7 and Figure 8 show the characteristics of training, obtained from the validation data during training for the left and right eyes, respectively. The solid line indicates the characteristics for the own network (5 layers CNN), the dashed line describes the VGG-19 network results.

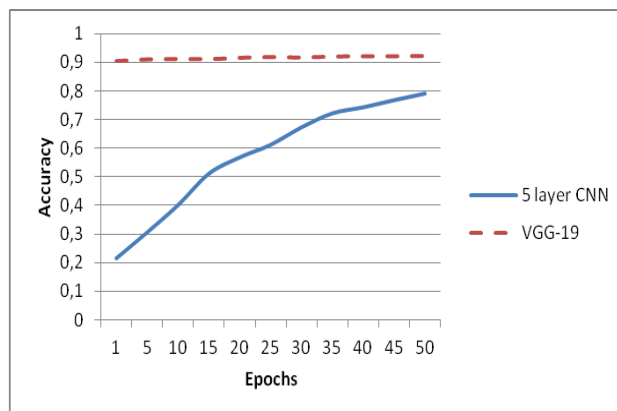


Figure 7. Characteristics of the validation set processing for the left eye

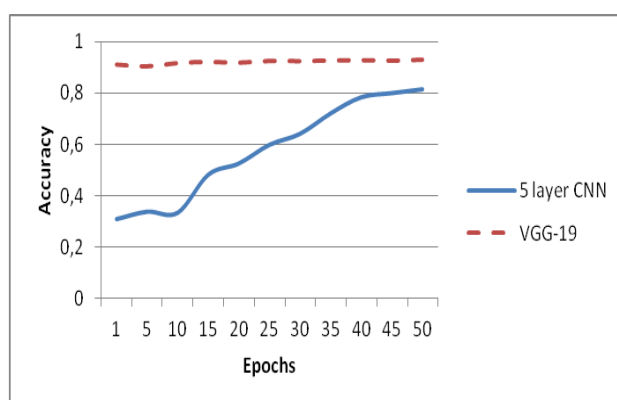


Figure 8. Characteristics of the validation set processing for the right eye

The analysis of the presented dependencies shows that it is advisable to use models trained using transfer learning, since the chosen architecture of the five-layer network is significantly inferior to the VGG-19 architecture. At the same time, as applied to the test sample, the characteristics of the share of correct recognitions have somewhat decreased. The results are presented in Table 1.

Network	Eye	Accuracy rate
VGG-19	Left	0.906
VGG-19	Right	0.911
5 Layer CNN	Left	0.792
5 Layer CNN	Right	0.811

Table 1. Percentage of correct recognition

Another important factor for the chosen task is the low level of errors of the first kind, when the decision is made that the eyes are open with the eyes closed. For a test sample of 2000 images, let us calculate recall by the formula (1):

$$recall = \frac{TP}{TP + FN}, \quad (1)$$

where TP is the number of correctly recognized images of closed eyes;

FN is the number of closed-eye images recognized as open-eyed images.

Table 2 lists the recall characteristics for both architectures and both eyes.

Network	Eye	Error rate
VGG-19	Left	0.093
VGG-19	Right	0.089
5 Layer CNN	Left	0.225
5 Layer CNN	Right	0.211

Table 2. Probability of skip target error

As can be seen from the Table 2, the probability of a skip target error for networks with the VGG-19 architecture is approximately 2 - 2.5 times less than for a network trained from scratch. However, a decrease in this indicator is required by an order of magnitude or even more.

5. RESULTS OF A VIDEO STREAM PROCESSING

Testing the operation of the algorithm shown in the diagram in Figure 5 was performed with a Bluesonic webcam (Full HD, IR illumination) in daytime conditions. In this case, the video sequence of images was processed with a slight delay. Figures 9-11 show examples of determining different eye conditions based on the VGG-19 model.

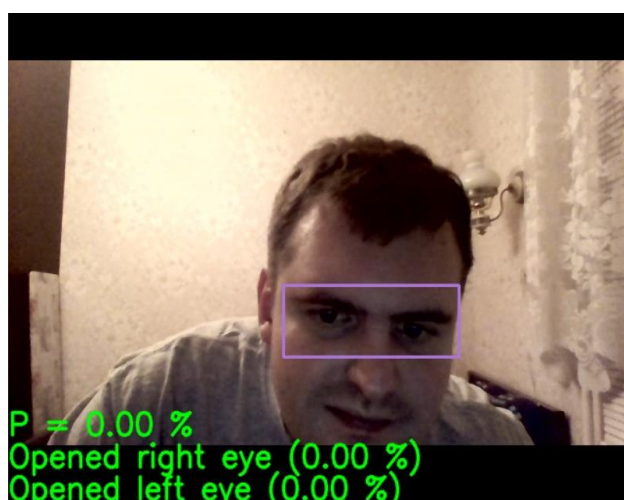


Figure 9. Eyes open

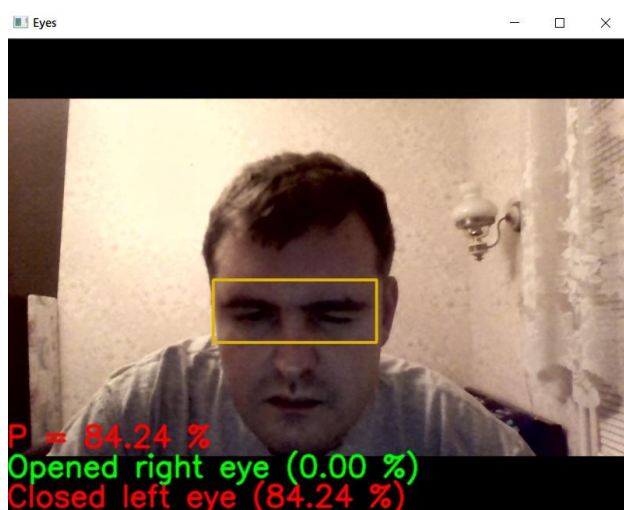


Figure 10. One eye closed

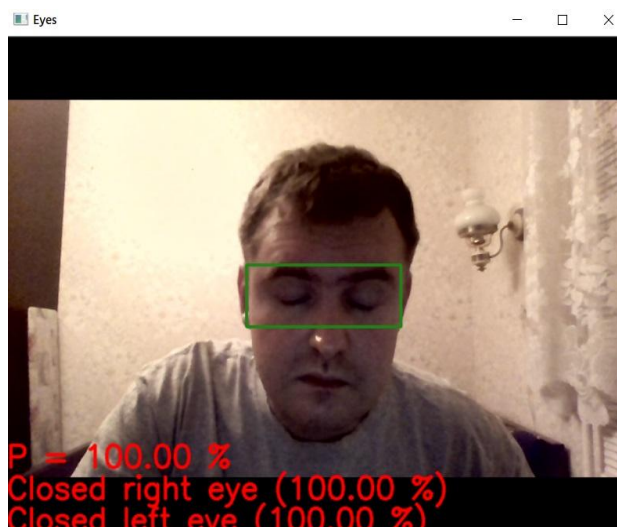


Figure 11. Eyes closed

However, the results of processing on the images from this camera are generally comparable to the results on the test sample using VGG-19 and are presented in Table 3. Here 200 images were taken, in which the share of correct recognition was calculated.

Network	Eye	Accuracy rate
VGG-19	Left	0.877
VGG-19	Right	0.869
5 Layer CNN	Left	0.704
5 Layer CNN	Right	0.733
SSD Mobile	Both	0.84
[14]		

Table 3. Proportion of correct recognition based on data from a video camera

Obviously, the VGG-19 model is more robust and better able to generalize data. Further improvement in performance can be achieved using data from a specific video camera during training.

6. CONCLUSION

In this paper, the main algorithms of computer vision for working with faces were considered, and several datasets with such images were presented. A comparative study was carried out on the results of recognizing the states of open and closed eyes on the face, which showed the need to use transfer learning when solving this problem. The obtained characteristics for a network based on the VGG-19 architecture are significantly superior to a simple five-layer convolutional neural network trained from scratch. The obtained characteristics in terms of the share of errors of the first kind (skip target) currently do not allow to unambiguously recommend the algorithm for practical application, however, due to the expansion of the training sample and the study of other architectures in the future, it seems to reduce the errors several times compared to the 9% presented in the work. The video processing algorithm also showed that the model based on the VGG-19 architecture remains stable with only a slight deterioration in the recognition quality in terms of accuracy.

ACKNOWLEDGEMENTS

This study was supported by a grant from the Russian Foundation for Basic Research, Project No. 19-29-09048.

REFERENCES

- Kamberova, N. 2017: 20% of all road accidents are associated with falling asleep while driving. [Electronic resource]. Access mode: <http://kazan-journal.ru/news/staraya-lenta-novostey/20-vseh-dtp-svyazany-s-zasyipaniem-za-rulem>, Access date: 10/01/2021.
- Andriyanov, N.A, Andriyanov, D.A. 2020: Pattern Recognition on Radar Images Using Augmentation. *Proceedings - 2020 Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology, USBEREIT 2020*, 289–291. Institute of Electrical and Electronics Engineers Inc. doi: 10.1109/USBEREIT48449.2020.9117669
- Wang, Sh., Yin, Sh., Hao, L., Liang, G. 2021: Multi-task face analyses through adversarial learning. *Pattern Recognition*, volume 114, 107837. doi: 10.1016/j.patcog.2021.107837.
- Choy, C., Lee, J., Ranftl, R., Park, J., Koltun, V. 2020: High-Dimensional Convolutional Networks for Geometric Pattern Recognition. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 11227-11236
- Andriyanov, N. A., Vasil'ev, K. K., Dement'ev, V. E. 2019: Investigation of filtering and objects detection algorithms for a multizone image sequence. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2-W12, vol. 42, 7–10. doi: 10.5194/isprs-archives-XLII-2-W12-7-2019
- Beal, J., Kim, E., Tzeng, E., Park, D.H., Zhai, A., Kislyuk, D. 2020: Toward Transformer-Based Object Detection [Electronic resource] Access mode: <https://arxiv.org/abs/2012.09958>, Access date: 13/02/2021
- Yu, H., Huang, Y., Pi, L., Zhang, Ch., Li, X., Wang, L. 2021: End-to-end video text detection with online tracking. *Pattern Recognition*, 113, 107791, doi: 10.1016/j.patcog.2020.107791.
- McDonald, A.D., Lee J.D., Schwarz, C., and Brown, T.L. 2018: A contextual and temporal algorithm for driver drowsiness detection. *Accident Analysis & Prevention*, 113, 25–37.
- Johnson, V. 1998: PERCLOS: A Valid Psychophysiological Measure of Alertness As Assessed by Psychomotor Vigilance. *USMSCAT*, 25, 10-17
- Wei, C.S., Wang, Y.T., Lin, C.T. and Jung, T.P. 2018: Toward Drowsiness Detection Using Non-hair-Bearing EEG-Based Brain-Computer Interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 1, 11-25
- Kartsch, V.J., Benatti, S., Schiavone, P.D., Rossi, D., and Benini, L. 2018: A sensor fusion approach for drowsiness detection in wearable ultra-low-power systems. *Information Fusion*, 43, 66–76
- Tateno, S., Guan, X., Cao, R., Qu, Z. 2018: Development of Drowsiness Detection System Based on Respiration Changes Using Heart Rate Monitoring. *57th Annual Conference of the Society of Instrument and Control Engineers of Japan, SICE 2018*, 1664–1669.
- Tayab, M.K., Anwar, H., Ullah, F., Rehman, A.U., Ullah, R., Iqbal, A., Lee, B.H., and Kwak, K.S. 2019: Smart Real-Time Video Surveillance Platform for Drowsiness Detection Based on Eyelid Closure. *Wireless Communications and Mobile Computing*, 2019, 1–9.
- Shakeel, M.F., Bajwa, N.A., Anwaar, A.M., Sohail, A., Khan, A., and Rashid, H. 2019: Detecting Driver Drowsiness in Real Time Through Deep Learning Based Object Detection. *Springer*, 283–296.
- Celona, L., Mammana, L., Bianco, S., and Schettini, R. 2018: A multi-task CNN framework for driver face monitoring. *IEEE International Conference on Consumer Electronics - Berlin, ICCE-Berlin*, 2018, 1–4.
- Kolomiets, V. 2014: Analysis of existing approaches to face recognition [Electronic resource] Access mode: <https://habr.com/ru/company/synesis/blog/238129/> Access date: 25/01/2021
- Kuznetsov, V. 2020: Allocation and face recognition. [Electronic resource] Access mode: <http://wiki.technicalvision.ru/index.php/> Access date: 26/01/2021
- Sun, J., Fu, Y., Li, S., He, J., Xu, C., Tan, L. 2018: Sequential human activity recognition based on deep convolutional network and extreme learning machine using wearable sensors. *Journal of Sensors*, 45-57
- Koonce B. 2021: VGG Network. *Convolutional Neural Networks with Swift for Tensorflow*, 35-50 doi: 10.1007/978-1-4842-6168-2_4