UDC 004.93
IRSTI 20.23.15

https://doi.org/10.55452/1998-6688-2026-23-3-282-297

[1]Smakanov B.S.,
PhD, ORCID ID: 0000-0002-4343-0463,
e-mail: s.bauyrzhan.10@gmail.com
[1*]Uvaliyeva I.M.,
PhD, Associate Professor, ORCID ID: 0000-0002-2117-5390,
e-mail: iuvalieva@ektu.kz

[1]D. Serikbayev East Kazakhstan Technical University, Ust-Kamenogorsk, Kazakhstan

MODELING DRIVER STATE MONITORING BASED ON A THEORETICAL AND EMPIRICAL APPROACH

Abstract

This article addresses the problem of improving face detection reliability in video streams used in driver state monitoring systems. Reliable localization of the driver's face is considered a preliminary computer vision stage required for the subsequent analysis of eye state, gaze direction, facial behavior, and fatigue-related visual cues. The purpose of this study is to develop and experimentally evaluate an adaptive face detection method that maintains stable performance under variations in illumination, head pose, image noise, and partial face occlusion while operating with a fixed, externally calibrated camera. The proposed method combines initial face-region detection using the Viola–Jones algorithm with SVM-based verification, external camera calibration, and incremental parameter updating based on high-confidence detections used as pseudo-labels. The method was evaluated using 18,000 video frames obtained from 12 recordings involving seven volunteers under different shooting conditions. The experimental results showed that the integrated method provided higher face detection performance than the standalone Viola–Jones algorithm and the Viola–Jones method combined with a conventional SVM classifier. The proposed method achieved a Precision of 0.91 and a Recall of 0.90 while maintaining a processing speed of 22 frames per second, which is sufficient for real-time operation. The study evaluates face detection and does not directly classify driver fatigue, drowsiness, eye state, or facial expressions. The obtained results demonstrate that the proposed method can be used as a lightweight preliminary component of driver state monitoring systems operating on embedded hardware with limited computational resources.

Keywords: video analytics, face detection, driver monitoring system, computer vision, video stream, Viola–Jones method, SVM classifier, adaptive detection, camera calibration, real-time processing.

Received July 21, 2026; revised August 28, 2026; accepted August 31, 2026.

Introduction

Studies of recent years show that a significant part of road accidents is associated with the physiological and psycho-emotional state of the driver [1–6]. In this regard, there is a growing interest in monitoring systems that automatically assess the driver's condition in the process of movement. Such systems allow you to detect signs of fatigue and decreased attention early and warn the driver in advance.

The basis of modern driver monitoring systems is computer vision technologies. According to the video frames, the driver's face, the state of his eyes and the position of his head in space are analyzed. In the course of the experiment, it was noted that the quality of detection is significantly influenced by the following factors:

- change in lighting level;
- turn and tilt of the head;
- change in the driver-to-camera distance and the corresponding face scale in the frame;

- video noise;
- partial closure of the face.

When the light is unstable or the driver turns his head, the detection accuracy of classical computer vision algorithms decreases and the number of erroneous determinations increases [7].

In studies to solve this problem, along with classic computer vision methods, YOLO, RetinaFace, MTCNN and other deep learning models have been widely used. Nevertheless, the adaptation of classical algorithms for systems with limited computing resources still remains relevant. Research and literary data show that the Viola - Jones algorithm processes video frames at high speed. Nevertheless, when lighting conditions change or the angle of rotation of the driver's head relative to the camera increases, the detection accuracy of the Viola–Jones algorithm decreases [8–10]. Although deep learning models show much higher results in such complex situations, their efficient operation requires hardware platforms with higher computing power. In this regard, the use of such models in built-in automotive systems with limited computing capabilities is not always acceptable.

For the driver monitoring system, the detection accuracy alone is not enough. The algorithm should not only process the video stream in real time, but also work stably during changes in lighting, movement of the driver's head and other external influences. Therefore, in the course of the study, priority was given to the development of an approach that has a low computational load and can adapt to variable shooting conditions. Such a solution provides reliable face localization under variable shooting conditions and can serve as an input stage for subsequent driver-state analysis.

Before the experiment, an external calibration of the camera was carried out. The position of the camera in space relative to the driver was calculated, and based on these parameters, a coordinate system was formed. These parameters were then used to calculate the position of the driver's face in space. The calibration results have increased detection stability, especially in cases where the driver's head is turned or tilted.

The working capacity of the model was tested under different shooting conditions. The following five experimental factors were evaluated: the illumination level; the driver’s head rotation and tilt angles; the driver-to-camera distance and the corresponding face scale in the frame; partial face occlusion; and the level of video noise. The camera position and orientation remained constant throughout all experiments.

This organization made it possible to evaluate the influence of each factor on the algorithm separately. After the end of each test, the number of detected faces, false determinations and the time spent processing one frame were recorded. The results were then compared and the impact of each parameter on detection accuracy and processing performance was analyzed.

The experiments carried out have confirmed that the proposed approach works effectively. The use of the Viola–Jones algorithm in conjunction with the SVM classifier and external camera calibration results reduced the number of erroneous determinations and increased the reliability of the system's operation in complex shooting conditions. During the test, the algorithm retained the processing speed in real time. The results obtained showed that this approach can be used in driver monitoring systems.

Materials and methods

During the experiment, the video camera was rigidly mounted inside the vehicle so that the driver’s face was fully visible in the frame. The position and orientation of the camera remained unchanged throughout the entire data collection process. External camera calibration was performed once before recording to determine the fixed spatial relationship between the camera and the driver. The experimental variations concerned the shooting conditions and the driver’s position relative to the fixed camera, rather than changes in the camera position itself.

To assess the effectiveness of the algorithm, a data set was formed from 12 video recordings with a total duration of about three hours. About 18,000 frames were taken from the videos, of which 12,000 were used for training, and the remaining 6,000 were used for testing the model. All

performance indicators presented in Tables 1 and 2 were calculated exclusively using the independent test subset consisting of 6,000 frames. The data was collected in such a way as to cover different face turning angles and partial closure scenarios, both in daytime and evening light conditions. The videos were filmed with the participation of the authors in real transport conditions. The generated data sets are not included in open databases and were used only as part of this study.

The study involved seven volunteers, five of whom were men and two were women. The age of the participants ranged from 21 to 46 years. Participants wearing glasses and with beards or mustaches were also included to ensure the diversity of the data set. The videos were filmed both in natural light and in an environment where there was not enough light. At the same time, different head turning positions in relation to the camera were deliberately introduced, and the ability of the algorithm to work in variable shooting conditions was assessed.

12,000 frames were selected from the generated dataset to train the SVM classifier. The remaining 6,000 frames were used to evaluate and test the model. During the experiment, five factors were taken into account: the illumination level; the driver's head rotation and tilt angles; the driver-to-camera distance and the corresponding face scale in the frame; partial face occlusion; and the level of video noise. These data allowed us to assess the stability of the system in real conditions. Examples of frames taken from the image of the driver's face are shown in Figure 1.

Figure 1 – Examples of driver-face frames extracted from the video recordings

During the experiment, video recordings of the driver's face were used as the main research material. First, the videos were divided into individual frames and their quality was checked. Further experiments continued under different shooting conditions. During the experiment, the shooting conditions were deliberately changed. The same five factors were varied according to the experimental protocol, while the camera position and orientation remained constant. After each test, the resulting video recordings were processed and the operation of the algorithm was evaluated by the number of detected faces, false determinations and the time spent processing one frame. These indicators made it possible to compare the effect of different shooting conditions on the detection accuracy and processing speed. Video frame processing was performed using the standard functions of the OpenCV library.

The search for the face area was carried out in several stages. First, the motion detection method was used and areas of interest (ROI) were marked, where potential objects were located. These areas were then scanned with a scalable rectangular window and checked for the presence of a face in each position. The Viola–Jones algorithm based on Haar-like features and cascade classifier was used as the main detection method. However, practice shows that its accuracy decreases when the lighting is unstable or the face looks blurry. Therefore, we have introduced an additional SVM classifier. It was pre-trained in data sets that took into account face orientation, light level and video noise, and allowed to reduce the number of false determinations.

Before the start of the experiment, the external calibration of the camera was performed. The position of the camera, the direction of view and the angles of rotation were determined, and these

parameters were used to match points in three-dimensional space with the image plane. In our practice, such calibration made it possible to more accurately determine the position of the face in space in cases where the driver's head is turned or tilted.

The results of the study showed that changes in lighting, turning of the driver's head and partial face coverage are the factors that most affect the quality of detection. Although the accuracy of the classic Viola–Jones algorithm in such conditions has decreased, the use of the SVM classifier has increased the stability of the detection results.

Under normal conditions, the system reliably detected the face. And in complex frames, sometimes the wrong areas were marked or the face was not found at all. An individual analysis of incorrectly processed frames was carried out. In most cases, such errors are caused by insufficient lighting, turning the driver's face relative to the camera, or a decrease in video quality.

Figure 2 – Examples of challenging conditions: low illumination, changes in face orientation, and low image resolution

Part of the results obtained during the experiment is presented in Figure 2. Tests have shown that the work of the algorithm depends on the shooting conditions. Under normal lighting, the face was reliably detected, and when the light level decreased or the angle of rotation of the driver's head increased, the number of errors increased. While in some frames the face remains undetected, in some cases other objects similar to the face are incorrectly recognized.

To minimize these shortcomings, an SVM classifier was introduced into the algorithm. The SVM classifier re-evaluated the candidate regions marked by the Viola–Jones method and clarified their correspondence to the face. As a result, false determinations have decreased and the stability of the algorithm has improved.

Each video was first divided into separate frames, and then they were processed sequentially. All results were compared among themselves and the features of the algorithm's work in different shooting conditions were analyzed. The frames were then processed in turn and converted to a grayscale image. Then only those parts where movement was observed were selected, and the search for facial features was carried out in the same areas. Determining areas of interest in advance reduced unnecessary calculations and reduced the amount of data being processed. As a result, the processing time of one frame decreased, and the algorithm met the requirements for working in real time. Where t is the sequence number of the frame in the video stream.

The C_t matrix is formed as a result of disconnecting the movement zone from the original A_t matrix. Thus, its dimensions are of a limited nature: the width satisfies the conditions $w_C \leq m$, and the height satisfies the conditions $h_C \leq n$.

It is necessary to form a set of rectangular areas containing human faces within a given area of movement (1):

$$Face_i = \{x_i, y_i, w_i, h_i\} \quad (1)$$

For each rectangular area in the list, the coordinates of its center (x_i, y_i), as well as the width (w_i) and height (h_i) parameters are determined.

To identify the object, the image is analyzed using the full scan method. For this purpose, a search window with a width of W and a height of H is used. The window is first placed in the upper left corner of the image and covers the entire image in a horizontal direction, moving by one pixel with each step. After that, the same search is performed in the vertical direction. The size of the search window is gradually changed to identify objects of different sizes in the image. The general operation of the sliding-window scanning algorithm is illustrated in Figure 3.

In the classification process for determining the presence of a face image in the image area, class symbols of the type {-1,+1} are used for each position of the scan window, where -1 means «not a face», and +1 means the class «face». In the process of scanning the image, each location of the window is checked using a classification model. Based on it, the presence of a face image in the area under consideration is decided in accordance with predefined designations and classification rules.

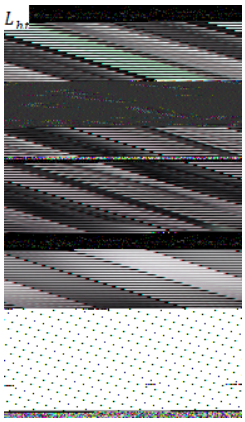

Figure 3 – Sliding-window scanning procedure
used for face detection

This approach allows us to consider the task of determining facial images in video frames as a two-class classification problem. Here, X represents the space of properties of objects, and Y represents a set of finite classes. The function F is considered as a decisive transformation that maps the characteristics of $x \in X$ to the corresponding $y \in Y$.

The function F defines the characteristics of object A by converting the elements in Set X to the D_f area, which forms the set of allowed values for a given object. In this case, the feature vector describing the object A and belonging to the complex X is based on the sign system $f_1, \dots, f_r$ [11–13] (2):

$$x = (f_1(a), \dots, f_r(a)) \tag{2}$$

In this context, objects are represented as feature vectors that describe their properties. In this case, the classifier must be able to accurately determine that any object belongs to a set X and reduce the probability of error for all possible X variants.

The effectiveness of the classifier is determined by its ability to reduce the probability of error for all possible inputs, which is the main criterion for the quality of its work.

When training the classifier F, a data set consisting of objects belonging to previously known Classes (3) is used:

$$D = \{(x_1, y_1), \dots, (x_m, y_m)\}, where\ y_i \in Y = \{-1; +1\} \tag{3}$$

In the proposed method, face detection is performed as a sequential two-stage procedure. First, the Viola–Jones algorithm processes each video frame and generates candidate face regions. Second, the SVM classifier verifies every candidate region and rejects detections that do not correspond to a face. The external camera calibration parameters are used to describe the fixed spatial relationship between the camera and the driver and to support the interpretation of pose-related features. This processing sequence is applied to every frame; the system does not switch between Viola–Jones and SVM. In the proposed configuration, the SVM parameters can additionally be updated using high-confidence pseudo-labeled samples obtained from the video stream.

When processing a video stream, the classifier settings were updated based on new frames. This allowed the model to adapt to changed shooting conditions and maintained the stability of the detection results. Examples of video frames obtained in normal and real road conditions are shown in Figure 4.

Figure 4 – Video frames obtained under (a) controlled and (b) real driving conditions

During the experiment, the camera position and orientation remained constant. Five experimental factors were evaluated: the illumination level; the driver's head rotation and tilt angles; the driver-to-camera distance and the corresponding face scale in the frame; partial face occlusion; and the level of video noise. The influence of these factors on face detection performance was analyzed under the same experimental protocol. The main features used in the study are as follows:

– α_h – angular deviation of the face along the X axis;

– α_v – angular deviation of the face in the vertical plane;

– α_z – the angle of inclination of the head relative to the frontal plane;

– L_{ht} – illumination level indicator;

– N_{os} – noise level in the image, where 0 means a perfect image, and 1 means a strongly distorted image that is not recognizable.

To unify the calculations, all parameters were normalized in advance to an interval of 0-1. The limit values were determined on the basis of experimental data. Since the face is not a completely symmetrical object, the angles of rotation and inclination were calculated at absolute values. Such normalization made it possible to directly compare the results obtained under different shooting conditions.

The quality of face detection was primarily dependent on the initial detection stage. Errors that occur at this stage also affect the result of the subsequent classification. Therefore, the influence of external factors on the characteristics of the face image was considered separately, and their relationship was described by Equation (4):

$$Q(a) = (\alpha_h, \alpha_v, \alpha_z, L_{ht}, N_{os}) \quad (4)$$

Each image in the training and test sets was attributed to a specific class, and a corresponding feature vector was created for it. This vector was used as input data in the later classification period. The formal description of the data set is given in Equation (5):

$$D = \{Q(a_1), Q(a_2), \dots, Q(a_n)\} = \{(\alpha_{h1}, \alpha_{v1}, \alpha_{z1}, Lht_1, Nos_1), \dots, (\alpha_{hn}, \alpha_{vn}, \alpha_{zn}, Lht_n, Nos_n)\} \quad (5)$$

At the next stage, a parametric assessment of the training set was performed (6):

$$D = \begin{Bmatrix} (\alpha_{h1}, \dots, \alpha_{hn}), (\alpha_{v1}, \dots, \alpha_{vn}), (\alpha_{z1}, \dots, \alpha_{zn}), \\ (Lht_1, \dots, Lht_n), (Nos_1, \dots, Nos_n) \end{Bmatrix} \quad (6)$$

Then a parametric analysis of the training set was carried out (6). The distribution of values for each sign was calculated, and the most common value was taken as mode. Based on it, a vector of generalized characteristics was obtained that characterize the training set. Its mathematical expression is given in Equation (7):

$$M_D = \{M\alpha_h, M\alpha_v, M\alpha_z, M_{Lht}, M_{Nos}\} \quad (7)$$

where $M\alpha_h, M\alpha_v, M\alpha_z, M_{Lht}, M_{Nos}$ - each parameter on mode.

The MD vector summarizes the characteristics that are often found in sample images [11]. The value of each parameter varies between predetermined minimum and maximum limits. This range is shown in Equation (8):

$$D_{min}^{max} = \begin{Bmatrix} (min_{a_h}, max_{a_h}), (min_{a_v}, max_{a_v}), \\ (min_{a_z}, max_{a_z}), (min_{Lht}, max_{Lht}), \\ (min_{Nos}, max_{Nos}) \end{Bmatrix} \quad (8)$$

To ensure effective classification, the characteristics of the images used during the testing phase must be within the permissible range defined by $D_{\min}$. The experimental results showed that face detection accuracy depended on the diversity of the training and test data. Changes in illumination and in the position of the driver's face relative to the fixed camera increased the number of false detections. Therefore, the training dataset included video frames obtained under different illumination conditions, head poses, driver-to-camera distances, and face positions, while the camera itself remained fixed.

The effectiveness of the SVM classifier was evaluated individually. The analysis showed that its performance depends not only on the size of the training set, but also on the variety of its composition. In the training set, where images with different levels of illumination, shooting angle and face position were included, the classification results were more stable. For this reason, when collecting video materials, the external parameters of the camera were determined in advance and special attention was paid to the variety of shooting conditions. Later, these materials were used to train the classifier.

In the process of evaluating the proposed method, along with the detection accuracy, the processing speed was also taken into account. Although convolutional neural networks were able to show high accuracy, the requirement they placed on computing resources was quite high. For the hardware platform used in the experiment, the use of such models did not allow stable operation in real time. For this reason, it was decided that the use of classical computer vision methods is more effective.

In the proposed model, the Viola–Jones algorithm performs the initial face detection stage, and the SVM classifier re-checks the identified areas. Such a division made it possible to use the strengths of each algorithm. The first provided processing speed, while the second reduced the number of erroneous determinations in complex frames.

During the experiment, only one algorithm could not give the same result in all cases. Therefore, the tasks of searching for a face and checking it were divided into two methods. First, Viola-Jones outlined possible facial areas. Later, the same areas were re-evaluated by the SVM. Such an approach did not require much additional calculation time, but the result of the determination was more stable. The parameter vector Q(a) was calculated for each processed frame to characterize the current

illumination, head pose, and image-noise conditions. This vector was not used to switch between the two algorithms. The Viola–Jones algorithm remained the candidate-region detector, while the SVM remained the verification stage under all experimental conditions. When the predefined update conditions were satisfied, high-confidence verified samples were used to incrementally update the SVM parameters.

In the course of the work, no approach was used to fully retrain the model for each new data set. Instead, the SVM classifier was adjusted in stages based on feature vectors derived from the new video frames. This approach follows the incremental update principle and significantly reduces the time required to retrain the model. As a result, incremental updating required less computation than complete model retraining. However, the additional verification and updating procedures reduced the overall processing speed of the proposed method to 22 FPS.

Manual marking for marking new frames was not carried out. Instead, the high–reliability detection results were obtained from the original Viola-Jones algorithm. Such face areas were adopted as a pseudo-label and then used when updating the SVM classifier.

If the detection reliability is higher than the predetermined threshold value, the corresponding feature vectors were included in the training phase. And frames with insufficient reliability were not involved in the process of updating the model. Such sampling reduced the accumulation of false samples and made it possible to maintain the stability of the classifier.

During testing, the model parameters were not changed with each frame processed. The update was performed only if the detection result was unstable or after a certain number of frames were processed. This selective update strategy avoided retraining the classifier for every processed frame and limited the additional computational overhead. Nevertheless, the complete proposed pipeline operated at 22 FPS, compared with 24 FPS for Viola–Jones combined with a static SVM and 28 FPS for the standalone Viola–Jones algorithm.

The computational complexity of the update algorithm was rated as $O(n \cdot d)$. Here n is the number of new samples participating in the update, and d is the size of the feature vector. It was confirmed during the experiment that this approach requires a much smaller computational resource compared to full model retraining.

Before starting the experiment, the drivers who participated in the study were explained the purpose of the work and the order of execution. Each participant gave their consent to the use of videos in scientific research. Participation was completely voluntary.

In the processing of video materials, no data was used that allows you to identify an individual. To analyze the results of the study, only the face area and the information necessary for the operation of the algorithm were used. Therefore, the personal data of the participants were not included in the published materials.

The collected videos were stored separately, and only the participants who participated in the study had access to them. The data was not shared with other organizations or used for commercial purposes.

The images obtained during the experiment were used only to test face detection algorithms. When processing personal data, the current regulatory requirements were taken into account and the confidentiality of information was maintained.

Results and discussion

At this stage of the experiment, the external parameters of the camera were determined. To determine the external parameters of the camera, reference objects were used, the position in space and geometric dimensions of which are known in advance [17-20]. As a result of the analysis of the video recordings, the position and angles of rotation of the camera were calculated. Later, the obtained parameters were used to more accurately determine the position of the face in space and estimate its position.

The CCTV camera vision system consists of the following main components:

1) the image structure is presented in the form of a rectangular pixel grid consisting of R rows and C columns.

2) the camera works in a local coordinate system formed by the X_c, Y_c, Z_c axes.

3) the camera is considered in the world (global) coordinate system,characterized by the axes X_w, Y_w, Z_w.

Samples of video frames obtained by the camera are shown in Figure 5.

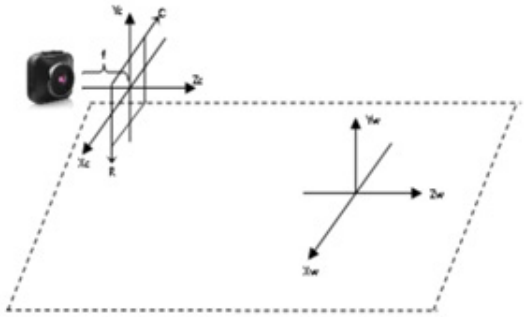

Figure 5 – Examples of driver-face frames captured by the fixed camera

The camera is characterized by an Optical Center Point (0, - f, 0) in the coordinate system, where f is the focal length. According to the camera matrix, the Rc plane is located in the $X_c Y_c$ plane, and the Z_c axis indicates the optical direction of the camera [20].

The camera calibration process is based on determining the correspondence between points in the image plane and the corresponding three-dimensional points in space. To solve this problem, the C_w^1 model is used, which describes the relationship between three-dimensional points in the global coordinate system and their projections into the image. This relationship is expressed taking into account the arbitrary scaling factor S (9).

$$\begin{bmatrix} s*r \\ s*c \\ s \end{bmatrix} = C_w^1 \begin{bmatrix} x^w \\ y^w \\ z^w \\ 1 \end{bmatrix} \tag{9}$$

The external parameters of the camera determine its position and orientation in space in the world coordinate system. They consist of two main groups:

- translation vector (10):

$$\mathrm{t} = [t_x t_y t_z]^T \tag{10}$$

- rotation matrix (11):

$$\mathrm{R} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & 0 \\ r_{21} & r_{22} & r_{23} & 0 \\ r_{31} & r_{32} & r_{33} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{11}$$

The translation vector gives the position of the camera in the world coordinate system, and the rotation matrix gives its orientation in space, that is, the direction of view. In turn, the matrix R is expressed as the product of three elementary rotations performed along the coordinate axes (12):

$$R = R_{ox} R_{oy} R_{oz}$$

$$R_{ox} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & cos\alpha & -sin\alpha & 0 \\ 0 & sin\alpha & cos\alpha & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, R_{oy} = \begin{bmatrix} cos\beta & 0 & sin\beta & 0 \\ 0 & 1 & 0 & 0 \\ -sin\beta & 0 & cos\beta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (12)$$

$$R_{oz} = \begin{bmatrix} cos\gamma & -sin\gamma & 0 & 0 \\ sin\gamma & cos\gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The R_{ox} matrix describes the rotation around the O_x axis at an angle α. Similarly, the R_{oy} matrix describes the rotation around the O_y axis at an angle β, and the R_{oz} matrix describes the rotation around the O_z axis at an angle γ. Accordingly, the values α, β and γ determine the angles of rotation performed around the axes O_x, O_y , O_z , respectively. These parameters describe the degree of rotation of each axis in three-dimensional space.

The transition from the global coordinate system to the camera coordinate system is carried out only on the basis of external parameters and reflects the change in the position and direction of the camera in external space (13):

$$T_W^C = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} x^c \\ y^c \\ z^c \\ 1 \end{bmatrix} = T_W^C \begin{bmatrix} x^w \\ y^w \\ z^w \\ 1 \end{bmatrix} \quad (13)$$

The T_w^C conversion, which describes the transition from the global coordinate system to the camera coordinate system, is reversible, that is, it allows you to accurately restore the original global coordinates from the coordinates in the camera system (14):

$$T_C^w = (T_w^C)^{-1}. \quad (14)$$

The structure of the transformation matrix T_w^C is determined by creating a transformation (14). This process consists of several sequential transformation steps, each of which plays an important role in the formation of the camera model, and as a result, the C_w^1 (15) camera model is determined:

$$T_w^1 = T_c^1 T_w^c \quad (15)$$

The external parameters of the camera were determined using the calibration methods described in the works [21–28]. As a result of calibration, the position of the camera in space, the direction of view and the main shooting parameters were calculated.

To clearly distinguish the evaluated configurations, three face detection methods were compared under the same experimental conditions and using the same test data. The first configuration, referred to as «Viola–Jones,» used only the standard Haar cascade detector with fixed parameters. The second configuration, referred to as «Viola–Jones + SVM,» used the Viola–Jones algorithm to generate candidate face regions and a previously trained static SVM classifier to verify these regions. In this configuration, the SVM parameters were not updated during video processing, and external camera calibration parameters were not incorporated into the detection procedure. The third configuration, referred to as «the proposed method,» combined Viola–Jones candidate detection, SVM-based verification, external camera calibration, and incremental SVM updating based on high-confidence pseudo-labels obtained from the video stream. Thus, the proposed method differs from the conventional Viola–Jones + SVM configuration through the additional use of camera calibration and online learning.

The effectiveness of the method was assessed by Accuracy, Precision, Recall and the number of frames processed per second (FPS). The results are presented in Table 1.

The proposed method was compared with the classic Viola–Jones algorithm in the same video format and in the same experimental conditions.

During the experiment, the method was evaluated under five experimental factors: variations in illumination; driver head rotation and tilt; changes in the driver-to-camera distance and the corresponding face scale; partial face occlusion; and video noise. The camera remained fixed in all experimental scenarios.

For each case, the detection accuracy was calculated individually, and the obtained indicators were compared among themselves. The results showed that the presented model works stably even in complex shooting conditions and provides higher detection reliability compared to the classic Viola-Jones algorithm.

Table 1 – Comparison of the performance indicators of face detection methods

Method	Accuracy	Precision	Recall	FPS
Viola–Jones baseline	0.82	0.79	0.76	28
Viola–Jones + static SVM	0.89	0.87	0.85	24
Proposed method: Viola–Jones + calibrated features + online SVM	0.91	0.91	0.90	22

Based on the corrected confusion matrix, the proposed method achieved an Accuracy of 0.9067, a Precision of 0.9122, and a Recall of 0.9000. When rounded to two decimal places, these values correspond to 0.91, 0.91, and 0.90, respectively. These recalculated indicators are consistent with the corrected entries presented in Table 2.

The staged comparison presented in Table 1 demonstrates the effect of adding static SVM verification to the Viola–Jones baseline and the combined effect of incorporating external camera calibration and online SVM updating into the proposed method. The individual contributions of camera calibration and online learning were not evaluated separately and will be investigated in future ablation experiments.

At the end of the experiment, the results obtained were analyzed for several indicators. Along with quantitative indicators, the features of the algorithm for processing video materials were also analyzed. As a result of the comparison, there was a decrease in the number of false determinations and an increase in the share of correctly identified faces. A similar trend was observed in all shooting conditions provided for during the experiment.

Cases when the level of illumination changed were considered separately. Although the rotation of the driver's head reduced the accuracy of detection, the use of the SVM classifier made it possible to correctly identify the face in many frames. Especially this was clearly noticeable in video frames, where the light fell unevenly or part of the face remained in the shade.

The proposed method improved the face detection indicators but introduced additional computational overhead. The processing speed decreased from 28 FPS for the standalone Viola–Jones algorithm to 24 FPS for Viola–Jones combined with a static SVM and to 22 FPS for the proposed method. Thus, the complete proposed method was approximately 21.4% slower than the standalone Viola–Jones baseline and approximately 8.3% slower than the static Viola–Jones + SVM configuration. This reduction was not considered negligible; rather, it represents a trade-off between detection reliability and processing speed. Nevertheless, the achieved speed of 22 FPS allowed continuous real-time video processing in the tested system.

The error matrix presented in Table 2 characterizes the classification quality of the proposed method. The results showed a decrease in false positive and false negative determinations. Compared to the Viola-Jones algorithm, the presented model gave more stable results even when the lighting and shooting conditions changed.

Table 2 – Confusion matrix for the proposed face detection method evaluated on the 6,000-frame test set

Actual class	Face detected	Face not detected
Face present	2700	300
Face absent	260	2740

The efficiency of the algorithms was evaluated on the basis of a matrix of classification errors. The results presented in Table 2 showed a decrease in the number of false positive and false negative determinations in the proposed method. Errors were most often observed when there was a sharp change in lighting, a turn of the driver's head at a large angle relative to the camera, or partial closure of the face.

Additional experiments were conducted to evaluate the face detection method under the five experimental factors defined in the Materials and Methods section. These factors were varied according to the same experimental protocol, and their influence on face detection performance was analyzed. The position and orientation of the camera remained unchanged. The results obtained for each case were compared and the influence of these factors on the accuracy of determination was analyzed.

In order to reduce processing time, zones where movement was observed were marked in advance. The face was not searched throughout the frame, but only in the areas of interest. Such a solution reduced the amount of data being processed and reduced the processing time per frame. Thanks to this, the system met the requirements for working in real time.

At the initial stage, only the possibilities of the Viola–Jones algorithm were studied. Under normal shooting conditions, it showed a high processing speed, however, when there was not enough light or the turning angle of the driver's head increased, the detection accuracy decreased. To reduce these limitations, an SVM classifier was introduced into the system and used in conjunction with the external calibration results of the camera. As a result, the quality of detection in complex frames has improved and the number of false determinations has decreased.

The tests carried out supplemented the information given in the literature. The use of camera calibration parameters in conjunction with the SVM classifier increased the algorithm's operational stability and improved the quality of face detection in complex situations.

The results of the experiment confirmed that the developed model works stably under different shooting conditions. The results obtained showed that this approach can be applied in real time in driver monitoring systems. There is also the possibility of using it in intelligent transport systems and other applications based on video analytics.

Conclusion

This study developed and experimentally evaluated an adaptive face detection method intended as a preliminary component of driver state monitoring systems. The proposed method combines the Viola–Jones algorithm for initial face-region detection with SVM-based verification, external camera calibration, and incremental parameter updating using high-confidence detections as pseudo-labels. This configuration was designed to maintain a low computational load while improving detection reliability under variable shooting conditions.

The experimental evaluation was conducted using 18,000 video frames obtained from 12 recordings involving seven volunteers. The tests covered the five experimental factors defined in the Materials and Methods section: illumination, driver head rotation and tilt, driver-to-camera distance and the corresponding face scale, partial face occlusion, and video noise. Compared with the standalone Viola–Jones algorithm and the Viola–Jones method combined with a conventional SVM classifier, the integrated method demonstrated more reliable face detection under complex conditions. It achieved a Precision of 0.91 and a Recall of 0.90 while processing 22 frames per

second. Although this processing speed was lower than that of the standalone Viola–Jones algorithm, it remained sufficient for real-time operation on hardware with limited computational resources.

The results demonstrate that the proposed method can provide reliable face localization as an initial stage of a broader driver state monitoring system. However, the present study evaluates face detection only and does not directly classify fatigue, drowsiness, eye state, gaze direction, or facial expressions. Therefore, the reported results should be interpreted as evidence of improved face detection performance rather than direct validation of driver fatigue recognition.

The limitations of the study include the relatively small number of participants, the use of a non-public dataset, and the limited range of experimental scenarios. Future research will focus on evaluation using larger and more diverse datasets, separate quantitative analysis of the contribution of camera calibration and online learning, and integration of the proposed face detection method with modules for eye-state, gaze, head-pose, and fatigue analysis.

Funding

This research is funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP25795477).

REFERENCES

1 Povichanov, A.A. Obzor problematiki i perspektivnyh metodov intellektual'nogo analiza dannyh o sostoyanii voditelej transportnyh sredstv [Overview of Issues and Promising Methods of Intelligent Data Analysis on the State of Vehicle Drivers]. Ekonomika i kachestvo sistem svyazi, (4), 159–172 (2024). (in Russian).

2 Kirillova, E.S., and Serikov, S.A. Metody i sredstva kontrolya sostoyaniya voditelya avtomobilya [Methods and Means of Monitoring the State of a Car Driver]. Mezhdunarodnyj zhurnal gumanitarnyh i estestvennyh nauk, (3-2), 169–172 (2024). (in Russian).

3 Ahmetvaleeva, I.V. et al. Podhod k monitoringu sostoyaniya ustalosti voditelej transportnyh sredstv [An Approach to Monitoring the Drowsiness State of Vehicle Drivers]. Vestnik Tekhnologicheskogo universiteta, 26 (4), 58–62 (2023). (in Russian).

4 Katasyov, A.S., Katasyova, D.V., and Sibgatullin, A.A. Nejrosetevaya model' ocenki funkcional'nogo sostoyaniya voditelej v sistemah transportnoj bezopasnosti [Neural Network Model for Assessing the Functional State of Drivers in Transportation Safety Systems]. Elektronika, fotonika i kiberfizicheskie sistemy, 3 (1), 69–80 (2023). (in Russian).

5 Bryzgalov, V.I., Karpushko, M.O., and Burgonutdinov, A.M. Prognoz kolichestva dorozhno-transportnyh proisshestvij na platnyh avtomobil'nyh dorogah [Forecast of the Number of Traffic Accidents on Toll Roads]. Vestnik Permskogo nacional'nogo issledovatel'skogo politekhnicheskogo universiteta. Prikladnaya ekologiya. Urbanistika, (2), 24–36 (2023). (in Russian).

6 Selina, A.D., and Venecianskij, A.S. Znachenie monitoringa zdorov'ya voditelya transportnogo sredstva [The Importance of Monitoring the Health of a Vehicle Driver]. Izmerenie. Monitoring. Upravlenie. Kontrol', (1), 65–71 (2024). (in Russian).

7 YUdin, D.A., and Mahon', YA.S. Vliyanie stressa i ustalosti voditelya kak determiniruyushchih faktorov dorozhno-transportnyh proisshestvij: analiz po dannym Rossijskoj Federacii [The Influence of Driver Stress and Fatigue as Determining Factors of Traffic Accidents: Analysis Based on Russian Federation Data]. Intellektual'nyj kompas: napravlenie nauchnyh otkrytij: sbornik, 95 (2026). (in Russian).

8 SHarova, D.E., and Garbuk, S.V. Metody ocenki kachestva sistem komp'yuternogo zreniya: evolyuciya podhodov, tendencii i ogranicheniya [Methods for Assessing the Quality of Computer Vision Systems: Evolution of Approaches, Trends and Limitations]. Informacionno-ekonomicheskie aspekty standartizacii i tekhnicheskogo regulirovaniya, (1), 35–41 (2026). (in Russian).

9 Boboqulov, S.R. et al. UDK 004.85 (075.8): Ispol'zovanie algoritma Non-Maximum Suppression dlya povysheniya tochnosti i skorosti obrabotki raspoznavaniya lic v real'nom vremeni [Using the Non-Maximum Suppression Algorithm to Increase the Accuracy and Speed of Real-Time Face Recognition Processing]. Innovatsion Texnologiyalar, 59 (3), 111–115 (2025). (in Russian).

10 Lohvickij, V.A., YAkovlev, E.L., and Bushev, I.V. Prakticheskoe sravnenie metodov komp'yuternogo zreniya i glubokogo obucheniya v zadache binarnoj klassifikacii izobrazhenij [Practical Comparison of

Computer Vision and Deep Learning Methods in the Binary Image Classification Task]. Intellektual'nye tekhnologii na transporte, (4), 89–98 (2025). (in Russian).

11 Shvets, O., Smakanov, B., Kovacs, L., and Gyorok, G. Intelligent system for driver support using two classifiers for simulation. Journal of Theoretical and Applied Information Technology, 100 (15), 4767–4782 (2022).

12 Chao, Q. et al. A survey on visual traffic simulation: Models, evaluations, and applications in autonomous driving. Computer Graphics Forum, 39 (1), 287–308 (2020). https://doi.org/10.1111/cgf.13803

13 Gao, M., and Shi, G.Y. Ship spatiotemporal key feature point online extraction based on AIS multi-sensor data using an improved sliding window algorithm. Sensors, 19 (12), 2706 (2019). https://doi.org/10.3390/s19122706

14 Dmitriev, E.A. Metod Violy-Dzhonsa [Viola-Jones Method]. Nauchnye issledovaniya i razrabotki studentov: materialy VI Mezhdunar. studench. nauch.-prakt. konf., 67–69 (2018). (in Russian).

15 Shvets, O., Smakanov, B., and Soltanbekov, S. The use of neuroanalytics in modern video surveillance systems. Multidisciplinary Academic Notes, Science Research and Practices: Proceedings of the XV International Scientific and Practical Conference, Madrid, Spain, 585–589 (2022). https://doi.org/10.46299/ISG.2022.1.15

16 Elharrouss, O., Almaadeed, N., and Al-Maadeed, S. A review of video surveillance systems. Journal of Visual Communication and Image Representation, 77, 103116 (2021). https://doi.org/10.1016/j.jvcir.2021.103116

17 Shvets, O., Smakanov, B., and Györök, G. Stabilization of environmental conditions to improve the performance of a mobile application for the state of the driver monitoring. Proceedings of the 17th International Symposium on Applied Informatics and Related Areas (AIS 2022), Obuda University, Székesfehérvár, Hungary, 98 (2022).

18 Zhang, Z. A flexible new technique for camera calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22 (11), 1330–1334 (2000). https://doi.org/10.1109/34.888718

19 Shvets, O., Smakanov, B., and Bakatbayeva, Zh. Monitoring of the state of a person at hazardous work. Proceedings of the 16th International Symposium on Applied Informatics and Related Areas, Obuda University, Székesfehérvár, Hungary, 88–91 (2021).

20 Ehsani, J.P. et al. Developing and testing a hazard prediction task for novice drivers: A novel application of naturalistic driving videos. Journal of Safety Research, 73, 303–309 (2020). https://doi.org/10.1016/j.jsr.2020.03.010

21 Figueiras, P. et al. Real-time monitoring of road traffic using data stream mining. 2018 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC), 1–8 (2018). https://doi.org/10.1109/ICE.2018.8436271

22 Fan, Y. et al. Multiple obstacle detection for assistance driver system using deep neural networks. International Conference on Artificial Intelligence and Security (Cham: Springer International Publishing), pp. 501–513 (2019).

23 Al Haddad, C., and Antoniou, C. A data–information–knowledge cycle for modeling driving behavior. Transportation Research Part F: Traffic Psychology and Behaviour, 85, 83–102 (2022). https://doi.org/10.1016/j.trf.2021.12.017

24 Arvin, R., Kamrani, M., and Khattak, A.J. The role of pre-crash driving instability in contributing to crash intensity using naturalistic driving data. Accident Analysis & Prevention, 132, 105226 (2019). https://doi.org/10.1016/j.aap.2019.07.002

25 Bellini, P. et al. Real-time traffic estimation of unmonitored roads. 2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech), 935–942 (2018).

26 Hochin, T., Shinohara, Y., and Nishizaki, Y. Detection of driver's eye fixation on a moving target by using line fitting. 2019 6th International Conference on Computational Science/Intelligence and Applied Informatics (CSII), 13–18 (2019).

27 Kang, M.J., Kwon, O.H., and Park, S.H. Development of a crash risk prediction model using the k-nearest neighbor algorithm. Advanced Multimedia and Ubiquitous Engineering: MUE/FutureTech 2018 12 (Singapore: Springer), pp. 835–840 (2019). https://doi.org/10.1007/978-981-13-1328-8_109

28 Li, R. et al. Driver drowsiness behavior detection and analysis using vision-based multimodal features for driving safety. SAE Technical Paper Series, 1 (2020).

[1]Смақанов Б.С.,
PhD, ORCID ID: 0000-0002-4343-0463,
e-mail: s.bauyrzhan.10@gmail.com
[1*]Увалиева И.М.,
PhD, қауымдастырылған профессор, ORCID ID: 0000-0002-2117-5390,
e-mail: iuvalieva@ektu.kz

[1]Д. Серікбаев атындағы Шығыс Қазақстан техникалық университеті,
Өскемен қ., Қазақстан

ТЕОРИЯЛЫҚ-ЭМПИРИКАЛЫҚ ТӘСІЛ НЕГІЗІНДЕ ЖҮРГІЗУШІ МОНИТОРИНГ ЖҮЙЕСІНДЕ БЕТ-ӘЛПЕТТІ АНЫҚТАУДЫ МОДЕЛЬДЕУ

Аңдатпа

Мақалада жүргізушінің күйін мониторингтеу жүйелерінде қолданылатын бейнеағыннан бетті анықтау сенімділігін арттыру мәселесі қарастырылады. Жүргізуші бетінің сенімді локализациясы көздің күйін, көзқарас бағытын, бастың қалпын және шаршаудың көрнекі белгілерін кейіннен талдауға қажетті компьютерлік көрудің бастапқы кезеңі ретінде сипатталады. Зерттеудің мақсаты – қозғалмайтын, сыртқы параметрлері алдын ала калибрленген камера арқылы түсіру кезінде жарықтандыру жағдайы, бастың қалпы, кескіндегі шу деңгейі және беттің ішінара жабылу дәрежесі өзгерген жағдайда тұрақты жұмыс істейтін бейімделгіш бетті анықтау әдісін әзірлеу және оны эксперименттік тұрғыдан бағалау. Ұсынылған әдіс Виола–Джонс алгоритмі арқылы бет аймақтарын бастапқы анықтауды, анықталған аймақтарды SVM-классификаторының көмегімен тексеруді, камераны сыртқы калибрлеуді және сенімділігі жоғары анықтау нәтижелерін псевдобелгілер ретінде пайдалану арқылы параметрлерді инкрементті жаңартуды біріктіреді. Әдіс жеті еріктінің қатысуымен әртүрлі түсіру жағдайларында жазылған 12 бейнежазбадан алынған 18 000 кадр негізінде бағаланды. Эксперимент нәтижелері біріктірілген әдістің базалық Виола–Джонс алгоритмімен және оның статикалық SVM-классификаторымен толықтырылған нұсқасымен салыстырғанда бетті анықтаудың анағұрлым жоғары көрсеткіштерін қамтамасыз ететінін көрсетті. Ұсынылған әдіс секундына 22 кадр өңдеу жылдамдығында Precision көрсеткіші бойынша 0,91, ал Recall көрсеткіші бойынша 0,90 нәтижеге қол жеткізді. Зерттеу аясында тек бетті анықтау нәтижелері бағаланды; жүргізушінің шаршауын, ұйқышылдығын, көзінің күйін немесе мимикасын тікелей жіктеу жүргізілген жоқ. Алынған нәтижелер ұсынылған әдісті нақты уақыт режимінде жұмыс істейтін жүргізуші күйін мониторингтеу жүйесінің есептеу ресурстарын аз қажет ететін бастапқы компоненті ретінде қолдануға мүмкіндік беретінін көрсетеді.

Түйін сөздер: бейнеталдау, бетті анықтау, жүргізушіні мониторингтеу жүйесі, компьютерлік көру, бейнеағын, Виола–Джонс әдісі, SVM-классификаторы, бейімделетін анықтау, камераны калибрлеу, нақты уақыттағы өңдеу.

[1]Смақанов Б.С,
PhD, ORCID ID: 0000-0002-4343-0463,
e-mail: s.bauyrzhan.10@gmail.com
[1*]Увалиева И.М.,
PhD, ассоциированный профессор, ORCID ID: 0000-0002-2117-5390,
e-mail: iuvalieva@ektu.kz

[1]Восточно-Казахстанский технический университет им. Д. Серикбаева,
г. Усть-Каменогорск, Казахстан

МОДЕЛИРОВАНИЕ МОНИТОРИНГА СОСТОЯНИЯ ВОДИТЕЛЯ НА ОСНОВЕ ТЕОРЕТИКО-ЭМПИРИЧЕСКОГО ПОДХОДА

Аннотация

В статье рассматривается проблема повышения надежности обнаружения лица в видеопотоке, используемом в системах мониторинга состояния водителя. Надежная локализация лица водителя рассматривается

как предварительный этап компьютерного зрения, необходимый для последующего анализа состояния глаз, направления взгляда, положения головы и визуальных признаков усталости. Целью исследования является разработка и экспериментальная оценка адаптивного метода обнаружения лица, обеспечивающего устойчивую работу при изменении освещения, положения головы, уровня шума изображения и частичном закрытии лица в условиях использования неподвижной камеры с предварительно выполненной внешней калибровкой. Предложенный метод объединяет первоначальное обнаружение областей лица с помощью алгоритма Виолы–Джонса, проверку обнаруженных областей посредством SVM-классификатора, внешнюю калибровку камеры и инкрементное обновление параметров на основе результатов обнаружения с высокой степенью достоверности, используемых в качестве псевдометок. Метод оценивался на наборе из 18 000 видеокадров, полученных из 12 видеозаписей с участием семи добровольцев в различных условиях съемки. Экспериментальные результаты показали, что интегрированный метод обеспечивает более высокие показатели обнаружения лица по сравнению с базовым алгоритмом Виолы–Джонса и его комбинацией со статическим SVM-классификатором. Предложенный метод достиг Precision 0,91 и Recall 0,90 при скорости обработки 22 кадра в секунду. В исследовании оценивается только обнаружение лица; непосредственная классификация усталости, сонливости, состояния глаз или мимики водителя не выполнялась. Полученные результаты показывают, что предложенный метод может использоваться как предварительный малоресурсный компонент системы мониторинга состояния водителя, работающей в режиме реального времени.

Ключевые слова: видеоаналитика, обнаружение лица, система мониторинга водителя, компьютерное зрение, видеопоток, метод Виолы–Джонса, SVM-классификатор, адаптивное обнаружение, калибровка камеры, обработка в реальном времени.