

Monitoring Of The State Of A Person At Hazardous Work

Olga Shvets
IT and intelligent system
sub-department
East-Kazakhstan

Technical University
Ust-Kamenogorsk, Kazakhstan
olga.shvets75@gmail.kz

Bauyrzhan Smakanov
IT and intelligent system
sub-department
East-Kazakhstan

Technical University
Ust-Kamenogorsk, Kazakhstan
bauka.10.10@mail.ru

Zhanerke Bakatbayeva
IT and intelligent system
sub-department
East-Kazakhstan

Technical University
Ust-Kamenogorsk, Kazakhstan
zhanerke.25@mail.ru

Abstract—The relevance of increasing the level of safety of the population at work is obvious today. A special place is occupied by work associated with increased danger in the case of monotonous, routine work, when a person ceases to adequately control the situation due to fatigue or distraction of attention, for example, when guarding objects in front of monitors of security cameras, when the driver of a vehicle is driving for a long time, etc. The use of modern digital technologies in this area is a promising area of research. The results of the work are aimed at improving the safety of people engaged in monotonous activities in real time using neural networks. Application area is operational monitoring of the state of a person at hazardous work, for example security of objects or driver of a vehicle. In this study, the design of an intelligent system for monitoring a person's condition during hazardous work (awake or asleep) is considered.

Keywords—neural networks, emotions recognition, software development

I. INTRODUCTION

Currently, there is a rapid growth in the use of video surveillance systems, which is explained by a wide range of tasks solved by such systems and the constantly increasing availability of surveillance and communication facilities. Video surveillance systems are used in a wide variety of areas: monitoring of road transport systems, security and law enforcement, unmanned aircraft, control of production processes, as well as in many other areas. A person, his life and health are the highest value. The analysis of industrial accidents, injuries, accidents, as well as occupational diseases shows that the main reason for them is non-compliance with safety requirements, human ignorance of man-made hazards and methods of protection against them. The use of video surveillance systems allows you to analyse the behaviour of the observed objects, increase the efficiency of their control, the reliability and quality of the results obtained, and provide support in decision-making. With the development of computer methods for processing visual information, intelligent video surveillance systems capable of automatically analysing incoming information are becoming more and more attractive. Computerization of the process allows dramatically increasing the scale of monitoring and reducing the use of human resources, while increasing the reliability and impartiality of observation.

II. IMAGE PROCESSING FOR RECOGNIZING HUMAN EMOTIONS

There are two approaches to the recognition of human emotions based on the use of neural networks. The classical approach to the problem of classification of emotions is based on the classification of key points of the human face.

The location of key points fixes hard and non-rigid deformations of the face due to head movements and facial expressions. Algorithms such as PDM, CML, AAM, DPM, or CNN can be used to obtain key points of a human face [3, 6]. The next stage of recognition in the classical approach is the classification of key points. Support vector machines are well suited for the classification of key points. At the moment, the problem of finding key points is quite well studied and there are a large number of algorithms that allow you to obtain key points with accuracy sufficient for further classification by these points of human emotions. But, to use the classic approach, it is necessary that the position of the face in the image be aligned. An alternative to using the classical approach is the approach based on convolutional neural networks. A convolutional neural network is architecture of artificial neural networks aimed at efficient pattern recognition, aimed at the operation of calculating a new value of a selected pixel, taking into account the values of the surrounding pixels. Convolutional networks are a good basic solution for classifying various visual data, which was used in this work.

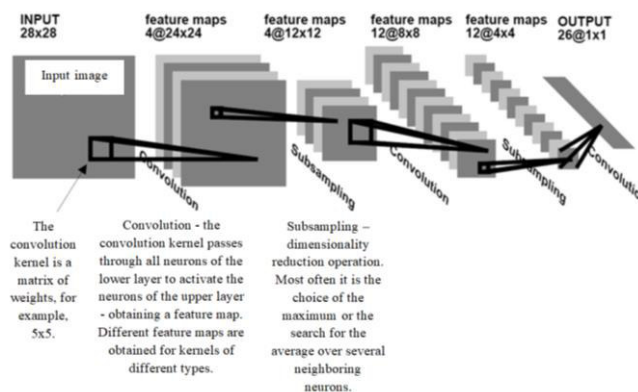


Fig. 1. Network architecture, Yann LeCun [6]

The convolution kernel, which is a square $n * n$ matrix, is used to calculate the value. When processing images, matrices of RGB-channels of pixels in rectangular coordinates are used as initial ones. Image filtering is one of the most fundamental operations of computer vision, pattern recognition, and image processing.

The convolutional neural network (CNN, LeNet) was introduced in 1998 by French researcher Yann LeCun as an extension of the neocognitron model [6] (see Figure 1).

The convolutional network model used consists of three types of layers:

- convolutional layers,

- subsampling layers and
- layers of a "normal" neural network - perceptron (see Figure 2).

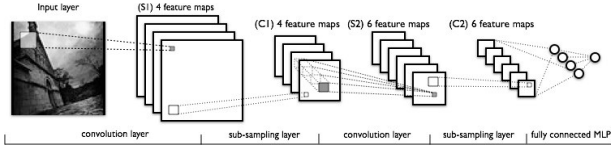


Fig. 2. Example of a Diagram of a convolutional network [6]

The first two types of layers form the input feature vector for the multilayer perceptron. The network can be trained using gradient methods.

The convolution operation can be described by the following formula [1, 4, 6].

$$(f * g)[m, n] = \sum_{k, l} f[m - k, n - l] \cdot g[k, l] \quad (1)$$

Here f is the original image matrix; g is the kernel (matrix) of the convolution.

Depending on the method of processing the edges of the original matrix, the result may be smaller than the original image, the same size or larger.

The convolutional layer implements the idea of local receptive fields, i.e. each output neuron is connected only to a certain (small) region of the input matrix and thus simulates some of the features of human vision.

In a simplified form, this layer can be described by the following formula.

$$x^l = f(x^{l-1} * k^l + b^l) \quad (2)$$

Here x is the output of layer l , $f()$ is the activation function, b is the shift coefficient, and the symbol $*$ denotes the operation of convolution of the input x with the kernel k .

At the same time, due to edge effects, the size of the original matrices is reduced.

$$x_j^l = f\left(\sum_i x_i^{l-1} * k_j^l + b_j^l\right) \quad (3)$$

Here x_j^l is the feature map j (output of layer l), $f()$ is the activation function, b_j is the shift coefficient for the feature map j , k_j is the kernel of convolution number j , x_i^{l-1} is the feature map of the previous layer.

Using a subsampling layer can improve the recognition of zoomed samples (reduced or enlarged). Layers of this type reduce the size of the input feature map (usually by a factor of 2). This can be done in different ways, in this program the method of selecting the maximum element (max-pooling) was used - the entire feature map is divided into cells of 2x2 elements, from which the maximum by value are selected. Formally, the layer can be described as follows.

$$x^l = f(a^l \cdot \text{subsample}(x^{l-1}) + b^l) \quad (4)$$

Here x is the output of the layer $l()$ is the activation function, a , b are the coefficients, $\text{subsample}()$ is the operation of sampling local maximum values.

The last of the types of layers is the layer of a "regular" multilayer perceptron (MLP), it can be described by the following relationship.

$$x_j^l = f\left(\sum_i x_i^{l-1} \cdot w_{ij}^{l-1} + b_j^{l-1}\right) \quad (5)$$

Here x^l is the output of layer l , $f()$ is the activation function, b is the shift coefficient, w is the matrix of weight coefficients.

Consider the topology of a 7-layer convolutional network, the order of which is described below.

- input layer - image matrix;
- convolutional layer - a set of matrices of the same type (feature maps);
- subsampling layer – the previous set of matrices reduced by 2 times;
- convolutional layer – the previous set of matrices is combined into one in accordance with the scheme for connecting layers, then a new set is generated;
- subsampling layer – the previous set of matrices reduced by 2 times;
- MLP layer – the previous set of matrices is expanded into a vector and processed as MLP;
- MLP layer (output).

In this case, neurons (feature maps) of the second subsampling layer and the third convolutional layer are selectively connected, i.e. according to the adjacency matrix, which is specified as a network parameter. For a network with the number of feature maps in the second layer 7 and 9 in the third layer, it may look like this (see Table 1).

TABLE I. AN EXAMPLE OF AN ADJACENCY MATRIX OF CONVOLUTIONAL LAYERS CONNECTIONS

N	1	2	3	4	5	6	7	8	9
1	1	0	0	0	0	1	1	1	0
2	1	1	0	0	0	0	1	1	1
3	1	1	1	0	0	0	0	1	1
4	0	1	1	1	0	0	0	1	1
5	0	0	1	1	1		0	0	1
6	0	0	0	1	1	1	0	0	0
7	0	0	0	0	1	1	1	0	0

Thus, each output map is formed by a partial sum of the results of the convolutions of the input maps, for each such partial sum its own set of convolution kernels.

In order to ensure the full flow of mental processes, a person must be in a state of wakefulness. Only in conditions of optimal wakefulness can a person [3, 5]:

- the best way to receive and process information;
- recall the necessary electoral communication systems;
- program activities;
- exercise control over activities;
- correct errors;
- maintain the chosen direction of activity.

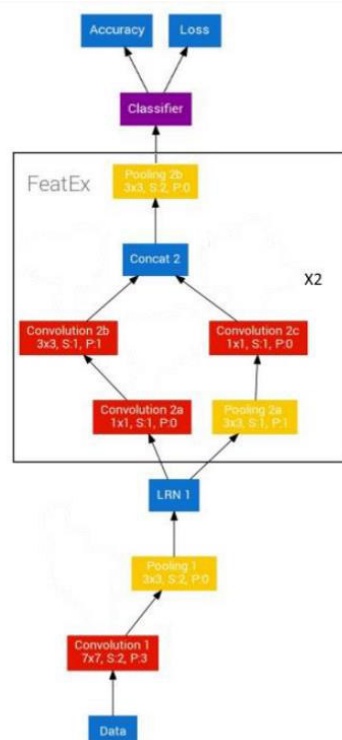


Fig. 3. Architecture of the model

There are the following levels of wakefulness: extreme level of tension, active wakefulness, and calm wakefulness.

The movements of the driver's head, describing its position at any moment in time, can be obtained using a static 2D image from the camera. In order to determine the position of the driver's head, which is not strictly parallel, but at a certain angle to the plane of the smartphone camera, Euler angles are calculated that describe the rotation of an object in three-dimensional Euclidean space. Angles of turning the head to the left or to the right, as well as tilting forward or backward, can be applied when a distraction of the driver's attention is detected, or a situation where the driver is "nodding", that is, it becomes difficult for him to keep his head in the normal position, respectively. When detecting the state of weakened attention of the driver, the determination of the position of his head relative to the body is used. In the first case, a situation is considered in which the driver's head should be directed directly in the direction of movement of the vehicle. The driver has identified signs of a weakened attention if his head is not looking at direction of vehicle movement for more than 1.5 seconds.

III. SOFTWARE

Let's start processing images (photographs) to identify a state other than wakefulness (sleep or falling asleep).

In the "Path to the folder with the source still frames of the video stream" field, specify the path to the folder with the source images. Then it is required to indicate a description of the person's states according to the figure, for example, a person is awake or a person is asleep (see Figure 4).

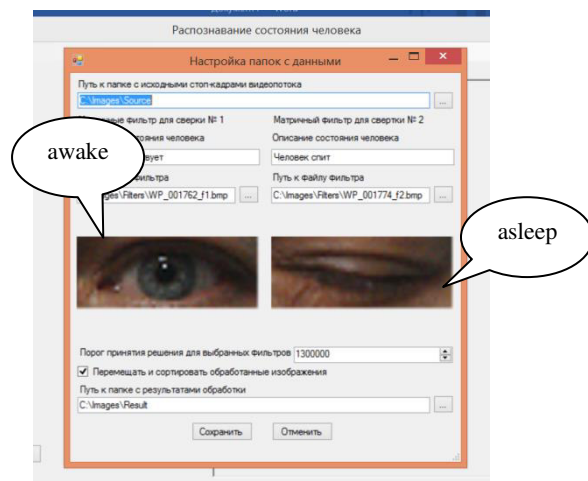


Fig. 4. Program settings window

Select filters from the Images \ Filters folder according to the picture. To do this, press the button "Start recognition" (see Figures 5, 6).

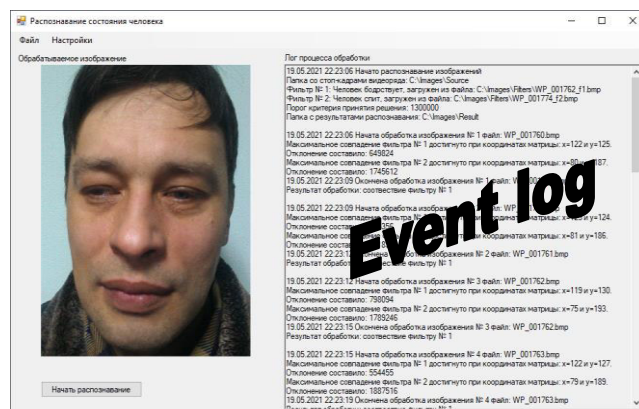


Fig. 5. Checking for compliance of the image with filter 1



Fig. 6. Checking for compliance of the image with filter 2

During the recognition process, the current recognized image is displayed in the left part of the window. In the right part of the window there is a log of the recognition process. The log displays information about the number of still frames of the video sequence, matches filter 1 (a person is awake) or filter 2 (a person is sleeping / falling asleep). If the freeze frame is not recognized as one of the filters, the message is displayed: "Failed to recognize the matching filters for <number> images."

14 test images are shown as an example: 8 with open eyes (matching filter 1), 4 with closed eyes (matching filter

2) and 2 with narrowed eyes (not matching filters) (see Figure 7).

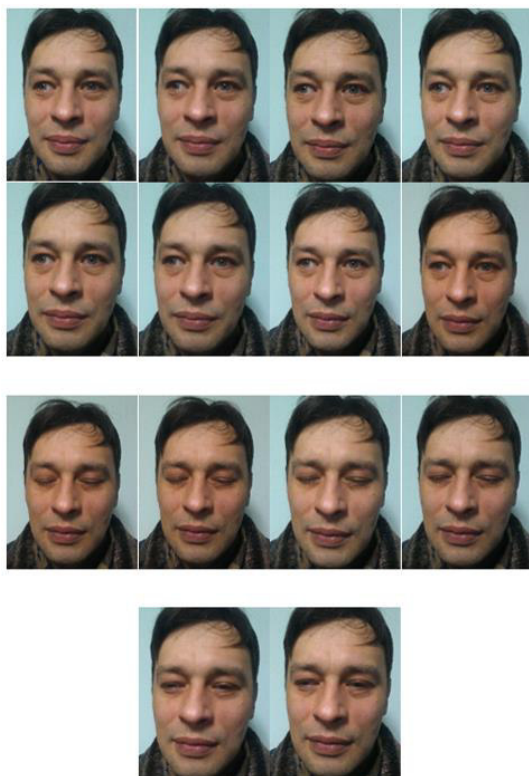


Fig. 7. Test images (photos)

The recognition accuracy of the test set of images is 100%. The program is customized for a specific person, which is also an additional advantage and a guarantee of the accuracy of state recognition. When changing an employee (driver, security guard), we can change the set of images, or add to the database.

IV. CONCLUSION

The monitoring of dangerous driver behaviour is based on the processing and analysis of driver images read from the smartphone camera in the vehicle cabin. It includes the detection of facial characteristics describing the condition and position of the eyes and mouth, the direction and movement of the driver's head.

It was developed intelligent system for monitoring the state of a person during hazardous work by Shvets O, Shokarev A., Bakatbayeva Zh. It was received a certificate on data registration in the state register of copyrighted objects of the Republic of Kazakhstan No. 18410, from June 04, 2021 (see Figure 8).

We are going to continue the research and to add more emotional states in the software, not only awake and asleep, but also extreme level of tension, active wakefulness, calm wakefulness, stress and fatigue.



Fig. 8. A certificate on data registration in the state register of copyrighted objects of the Republic of Kazakhstan No. 18410, from June 04, 2021 (Shvets O, Shokarev A., Bakatbayeva Zh.)

All the considered methods and algorithms are implemented in a software package in the C # programming language.

ACKNOWLEDGMENT

Part of the research was carried out with the support of the East-Kazakhstan technical university.

REFERENCES

- [1] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Ye Duan, O. Al-Shamma, J. M. Santamaria, M. A. Fadhel, M. Al-Amidie, L. Farhan, Review of deep learning: concepts, CNN architectures, challenges, applications, future directions, Journal of Big Data No. 8(1), March 2021, 74 p.
- [2] N. Xu, B. Price, S. Cohen, T. Huang, Deep image matting, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2970-2979.
- [3] R. Logan, G. B. Williams., M. Ferreira da Silva, A. Indani, N. Scholnicov, A. Ganguly, S. J. Miller, Deep Convolutional Neural Networks With Ensemble Learning and Generative Adversarial Networks for Alzheimer's Disease Image Data Classification, Frontiers in Aging Neuroscience, V. 13, 2021, p. 497, URL=<https://www.frontiersin.org/article/10.3389/fnagi.2021.720226>, DOI=10.3389/fnagi.2021.720226.
- [4] R. Patel, S. Patel, A Comprehensive Study of Applying Convolutional Neural Network for Computer Vision, International Journal of Advanced Science and Technology 6(6), 2020, pp. 2161-2174.
- [5] W. Xu, J. He, Y. Shu, H. Zheng, Advances in Convolutional Neural Networks, Advances and Applications in Deep Learning, Marco Antonio Aceves-Fernandez, IntechOpen, 2020, DOI: 10.5772/intechopen.93512, Available from: <https://www.intechopen.com/chapters/73604>.
- [6] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, Gradient-based learning applied to document recognition., Proceedings of the IEEE, november 1998, 86(11), pp. 2278-2324.