Implementation of the identification and recognition system cognitive behavior of the observed

A.V. Varuyeu, O.M. Demidenko, N.A. Aksionova, A.I. Kucharau
F. Skorina Gomel State University, Sovetskaya str, 104, Gomel, Belarus,
varuyeu@gmail.com, demidenko@gsu.by, nataliaksen@gmail.com, kucherov@gsu.by

Abstract*. This article describes and analyzes the development of a system for identifying and recognizing the cognitive behavior of students to determine interest in facial expressions. The purpose of the study is to find suitable technologies for the implementation of this system. The definition of emotions will allow organizing control over the quality of the educational process, conducting statistics on the cognitive behavior of students during classes and showing the level of interest of students in the material presented. The identification system will automatically determine and register the time of arrival and departure of students in real time. Based on the joint application of the Viola-Jones method and the nearest neighbors method using histograms of centrally symmetric local binary images, a system for face recognition in a real-time video sequence has been developed. The structure of the project is described and software is developed in the Python programming language using the Keras open source library. The developed system consists of two subsystems: an identification system and a cognitive behavior recognition system. The scientific novelty lies in an integrated approach to the development and research of algorithms for real-time face recognition and identification for solving applied problems.*

I. Introduction

Face recognition is a system designed to identify people in a static image or video. This technology has been around for decades. However, its use has become more visible in the last few years, as it is now used as part of innovative solutions, such as photo recognition of people and additional authentication on mobile devices.

By recognition we will understand the identification of an image of an unknown person with one of the well-known persons. Thus, the concept of "recognition" can be defined as the assignment of the object under study (face image), given as a set of observations, to one of the mutually exclusive classes (persons), or the conclusion that this object does not belong to known classes. The task of face recognition is relevant both in the field of intelligent environments and in security systems.

The authors of the article developed two Python applications using the *TensorFlow, OpenCV, Numpy, TKinter, Keras, MatPlotLib, PIL, Threading libraries*. These applications process the video stream (or selected file) from the webcam in real time. The Windows 10 operating system was used as a platform for work. *PyCharm* was used to work with the code.

The *Viola-Jones* method and convolutional neural network (*CNN*) are used for implementation. The *Viola-Jone*s method is considered one of the best in terms of the ratio of its speed to the accuracy of determination. Also, this method can successfully interact with other algorithms and can be adapted to specific needs and requirements. It works not only with static images, but can also process data in real time.

In the process of work, the authors faced the problem of the lack of a training set for each defined emotion: anger, joy, neutrality, sadness, surprise. To solve this issue, lectures were recorded on the subject of "System analysis and operations research" for third-year students. A database of students with defined emotions was created for the subsequent training of the system being developed.

II. Project structure

The process of face recognition is usually called a set of different tasks that serve to identify a person by a digital image or video fragment. In general, this process looks like this: after the system has received an image from the camera, the face boundaries (detection stage) are determined using algorithms. This is followed by the recognition stage, at which the face is transformed (its brightness

changes, it is aligned, scaled, etc.) and brought to a certain specified form. After that, the features are calculated and directly compared with the standards embedded in the database. This final stage of comparison is called identification or verification, depending on the system.

Processing of video stream frames consists of two main stages. At the first stage, the detection of faces in the frame takes place, and at the second - the recognition of the detected faces itself.

Figure 1 shows the interaction of the components of the developed identification subsystem. When the application is launched, it is first loaded from *CascadeClassifier files* using *OpenCV* and a ready-made model by the *Keras* module, which is imported from the *TensorFlow* library. The *OpenCV* searches for the webcam and starts recording, and also creates an application window. After that, every iteration of the inner loop, the frame is processed using *CascadeClassifier* and a face search is performed. After that, a frame showing the face is applied to the image. Then, through the *NumPy* library, the image is converted into a binary array, which is already fed to the finished model, and the result is output as text, as well as the processed image itself on the screen.

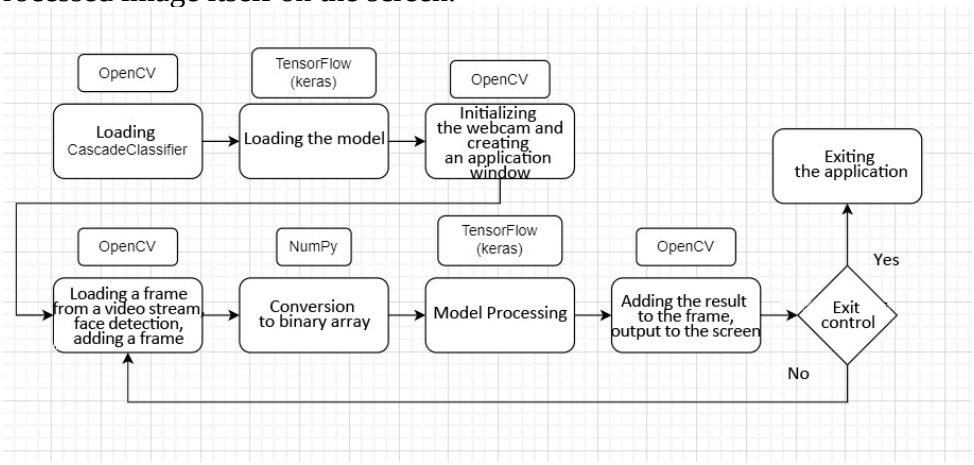

Fig. 1. Interactions of components of the identification subsystem

III. Creation of basic structures

The project consists of two classes: the main *App* class and the auxiliary *VideoCapture* class.

The *App* class is the main one; it initializes the *GUI*, widgets, main components and methods. First, let's create and describe all class variables:

the *class_labels* variable contains a tuple of all emotions that are defined in the program;

the variable *preds* is a list that contains the values of the emotions in the active frame and is updated as the program runs, the value of this variable is used to plot the active frame;

the *count* variable contains the number of frames that will be displayed on the emotion graph for count frames (Figure 2);

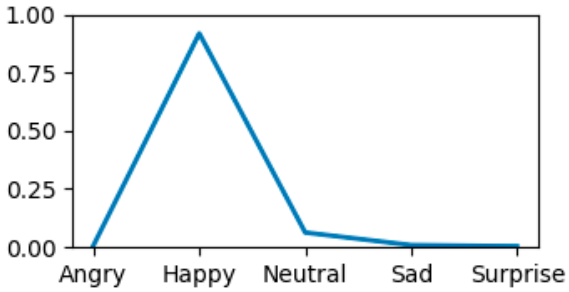

Fig. 2. Active cadre diagram

the file variable *file* contains the path and name of the file in which the result of the program is saved;

the *lines* variable is a list that contains data for count frames about each emotion, used to plot emotions for count frames (Figure 3);

the *face_classifier* variable contains a model for determining faces on a frame using the Viola-Johnson method;

the *classifier* variable contains a ready-made convolutional neural network model for determining emotions.

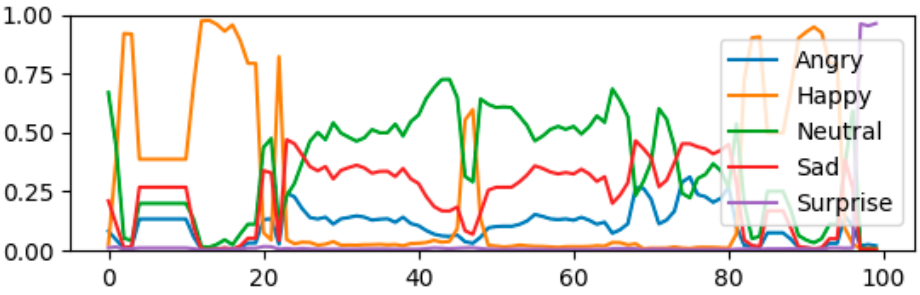

Fig.3. Diagram for "count" cadres

IV. Conclusion

As a result of the conducted testing, it was determined that the emotions "Neutrality", "Joy", "Sad" and "Surprise" are determined by the system virtually unmistakably. However, there are difficulties with defining the emotion "Anger". This problem is a consequence of the lack of training of the neural network for this emotion and requires further development. For the subsequent approbation of the system, the authors identified the following main testing criteria:

- the accuracy of the model in identifying the observed;
- model accuracy at different distances from the camera;
- model accuracy in different lighting conditions;
- the accuracy of the model at different positions and turns of the head;
- the accuracy of the model in determining a certain emotion;
- the accuracy of the model in determining emotions for several people in the frame;
- the accuracy of the model in determining emotions in different sexes.

As a result of the conducted research, suitable technologies have been developed and created for the implementation of an emotion identification and recognition system. The conducted testing showed 82% accuracy of identification of the observed. The accuracy of recognition of the five emotions obtained in this work is 65.56%. The results obtained show that the identification subsystem can be used to automate the registration of student attendance, automatically register the time of arrival and departure. The subsystem for determining the cognitive behavior of the observed requires further refinement.

References

[1] *P. Viola, M. Jones.* “Rapid Object Detection using a Boosted Cascade of Simple Features”, 2001.

[2] *V. Badrinarayanan, A. Kendall, R. Cipolla,* “A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation”.

[3] *Linda Shapiro, George Stockman*, Computer Vision. Binomial. *Knowledge Laboratory*, 2006, 752 p.

[4] *N. Aksionova, O. Demidenko, A. Varuyeu*. “Implementation of a system for determining students’ emotions by their facial expressions”. Bulletin of F. Skorina Gomel State University, 2022, No. 3 (132), pp. 82-87.