

ИНФОРМАТИКА

Известия Саратовского университета. Новая серия. Серия: Математика. Механика. Информатика. 2024. Т. 24, вып. 2. С. 275–286

Izvestiya of Saratov University. Mathematics. Mechanics. Informatics, 2024, vol. 24, iss. 2, pp. 275–286

<https://mmi.sgu.ru>

<https://doi.org/10.18500/1816-9791-2024-24-2-275-286>

EDN: ODMKXD

Article

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

O. M. Demidenko[✉], N. A. Aksionova, A. V. Varuyeu

Francisk Skorina Gomel State University, 104 Sovetskaya St., Gomel 246028, Belarus

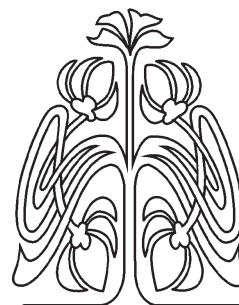
Oleg M. Demidenko, demidenko@gsu.by, <https://orcid.org/0000-0002-0601-0758>, AuthorID: 309622

Natalia A. Aksionova, nataliaksen@gmail.com, <https://orcid.org/0000-0002-1558-3064>, AuthorID: 1173355

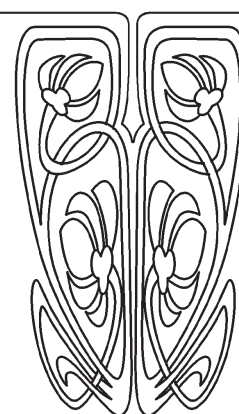
Andrei V. Varuyeu, varuyeu@gmail.com, <https://orcid.org/0000-0003-0235-0875>, AuthorID: 973250

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 the 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.

Keywords: face recognition, emotion detection, Viola – Jones method, computer vision, neural networks



Научный
отдел





For citation: Demidenko O. M., Aksionova N. A., Varuyeu A. V. Implementation of the identification and recognition system cognitive behavior of the observed. *Izvestiya of Saratov University. Mathematics. Mechanics. Informatics*, 2024, vol. 24, iss. 2, pp. 275–286. <https://doi.org/10.18500/1816-9791-2024-24-2-275-286>, EDN: ODMKXD

This is an open access article distributed under the terms of Creative Commons Attribution 4.0 International License (CC-BY 4.0)

Научная статья

УДК 004.932

Реализация системы идентификации и распознавания когнитивного поведения наблюдаемого

О. М. Демиденко[✉], Н. А. Аксёнова, А. В. Воруев

Гомельский государственный университет имени Франциска Скорины, Беларусь, 246028, г. Гомель, ул. Советская, д. 104

Демиденко Олег Михайлович, доктор технических наук, проректор по научной работе, demidenko@gsu.by, <https://orcid.org/0000-0002-0601-0758>, AuthorID: 309622

Аксёнова Наталья Андреевна, старший преподаватель кафедры автоматизированных систем обработки информации, nataliaksen@gmail.com, <https://orcid.org/0000-0002-1558-3064>, AuthorID: 1173355

Воруев Андрей Валерьевич, кандидат технических наук, заведующий кафедрой автоматизированных систем обработки информации, varuyeu@gmail.com, <https://orcid.org/0000-0003-0235-0875>, AuthorID: 973250

Аннотация. В статье описывается и анализируется разработка системы идентификации и распознавания когнитивного поведения учащихся для определения заинтересованности по мимике лица. Цель исследования — найти подходящие технологии для реализации данной системы. Определение эмоций позволит организовать контроль за качеством учебного процесса, провести статистику когнитивного поведения студентов во время проведения занятий и показать уровень заинтересованности обучаемых в излагаемом материале. Система идентификации позволит автоматически определять и регистрировать время прихода и ухода студентов в режиме реального времени. На основе совместного применения метода Виолы – Джонса и метода ближайших соседей с использованием гистограмм центрально-симметричных локальных бинарных образов разработана система распознавания лиц в видеопоследовательности в реальном времени. Описана структура проекта и разработано программное обеспечение на языке программирования Python с использованием библиотеки с открытым исходным кодом Keras. Разработанная система состоит из двух подсистем: идентификации и распознавания когнитивного поведения. Научная новизна заключается в комплексном подходе к разработке и исследованию алгоритмов распознавания и идентификации лиц в режиме реального времени для решения прикладных задач.

Ключевые слова: распознавание лиц, определение эмоций, метод Виолы – Джонса, компьютерное зрение, нейронные сети

Для цитирования: Demidenko O. M., Aksionova N. A., Varuyeu A. V. Implementation of the identification and recognition system cognitive behavior of the observed [Демиденко О. М., Аксёнова Н. А., Воруев А. В. Реализация системы идентификации и распознавания когнитивного поведения наблюдаемого] // Известия Саратовского университета. Новая серия. Серия: Математика. Механика. Информатика. 2024. Т. 24, вып. 2. С. 275–286. <https://doi.org/10.18500/1816-9791-2024-24-2-275-286>, EDN: ODMKXD

Статья опубликована на условиях лицензии Creative Commons Attribution 4.0 International (CC-BY 4.0)



Introduction

Recognition is the identification of the transformation of input information, which is appropriately defined by some features of recognizable objects in the output information, representing the conclusion about which class the recognizable object is found [1].

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 [2].

Face recognition has a number of advantages over other methods of human identification:

- no special or expensive equipment is required;
- physical contact with devices is not needed: it is enough just to pass or linger in front of the camera for a short time.

The authors of the article developed two Python applications using the *TensorFlow*, *OpenCV*, *Numpy*, *TKinter*, *Keras*, *Matplotlib*, *PIL*, and *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 – Jones method is considered one of the best in terms of the ratio of its speed to the accuracy of the 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, and surprise [3]. 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.

1. 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 [2]. 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.

The processing of video stream frames consists of two main stages. In the first stage, the detection of faces in the frame takes place, and in the second, the recognition of the detected faces itself [4].

When developing the identification subsystem, the Viola – Jones face detection method was used in this work, and the nearest neighbor method with the use of histograms of centrally symmetric local binary patterns was used for recognition.

As a result, the generalized algorithm for processing frames of the developed identification subsystem contains the following stages:

- 1) face detection;

- 2) processing of found faces using a Gauss filter;
 - 3) LBP transformation of the found faces with subsequent application of the mask of significant areas;
 - 4) calculation of histograms of found faces;
 - 5) Classification of persons by the nearest neighbor method by histograms [5].
- The general algorithm of the subsystem being developed is shown in Fig. 1.

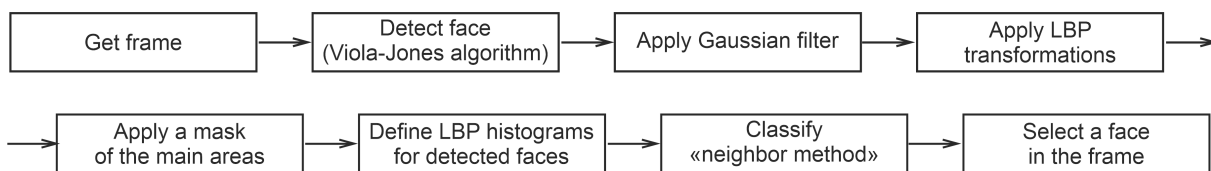


Fig. 1. Generalized algorithm for processing video stream frames

The Viola – Jones method is used to determine objects in real time. This algorithm was created to simplify the identification of faces in a photo or frame and is now widely used due to the processing speed because it allows you to recognize a face in a photo or frame in real time without using a large amount of computing power.

Convolutional neural network is used to determine emotions on a frame or photo because it has high accuracy, but it already requires large computing power, it is also processed longer.

The operation of a convolutional neural network can be interpreted as a transition from specific features of the image to more abstract, but numerous details, and then to more abstract, and even more numerous. Two main operations are defined in this neural network: the convolution operation and the subdiscretization operation.

The convolution operation is a passage of convolution kernels, or rather small matrices of weights of small size, through the layer, which eventually generates so-called prism maps. Each fragment from the feature map shows the presence of some feature that has been encoded into its convolution core.

The subsampling operation, also a subsampling operation, is used to reduce the dimension of the created feature maps. In this architecture, it is considered that the presence of a feature is more important than its location, so the maximum is usually selected from several neighboring neurons of the map. Due to this operation, the speed of the algorithm itself increases.

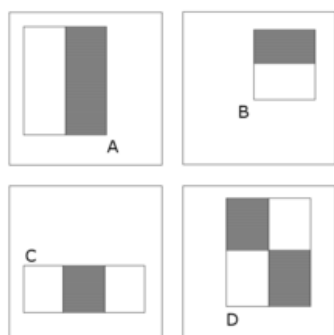


Fig. 2. Types of “features”

Viola – Jones algorithm was created in 2001 for real-time face recognition. It relies on the use of so-called “features”. The “features” used by the algorithm are based on the summation of pixels in rectangular areas (Fig. 2). The value of each “feature” is calculated as the difference between the sums from different parts of the “feature”. The use of these “features” is quite primitive, while they are sensitive to the vertical and horizontal features of the image and, as a result of all of the above, give a rather rough result.

Figure 3 shows a diagram of the use cases of the emotion recognition subsystem. This diagram shows a pattern of behavior and interaction between roles. The main block is the application itself. On the right is a user who has specific goals and behaviors regarding this application. The main role of the user is to launch the application and use the application. Launching the app involves checking for the presence of a camera. If there is a camera, then the application will work correctly. Otherwise, the application will not start. The application will also work correctly if the user’s face is in the field of view of the camera. Next, let’s take a closer look at the role of the developer.

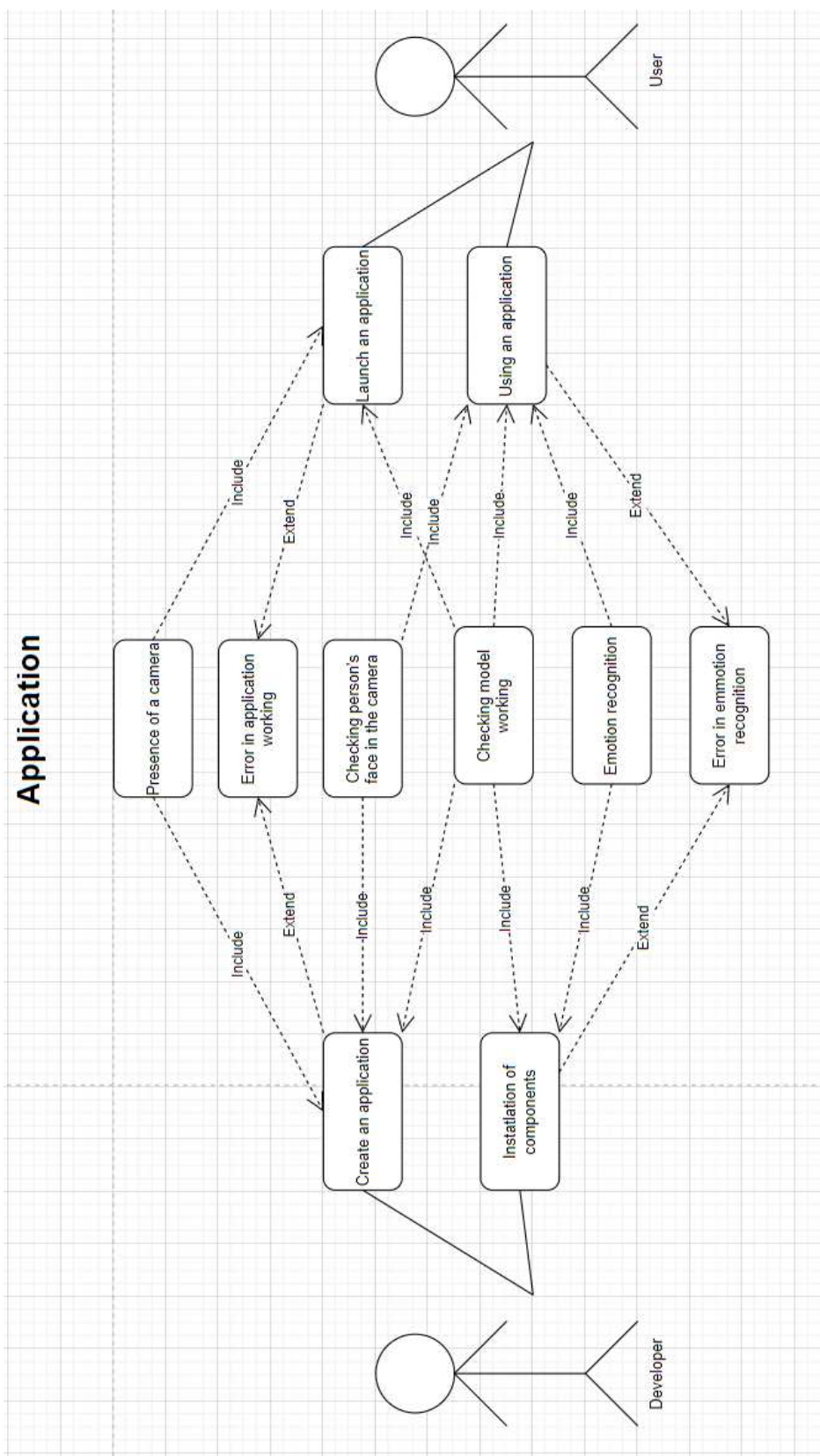


Fig. 3. A case diagram of the emotion recognition subsystem

The main task of the developer is to install all the necessary components and develop an application. The installation of components includes checking the operation of the model, if the model is installed and configured correctly and the necessary libraries are connected, then the application will recognize emotions without any problems. Otherwise, a recognition error will occur. Creating an app is an important process related to setting up a camera. Since all interaction with AR objects occurs through it, the camera must be connected and properly configured; otherwise, the application will not work and perform its original facial recognition function.

2. Creation of basic structures

Figure 4 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, for 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.

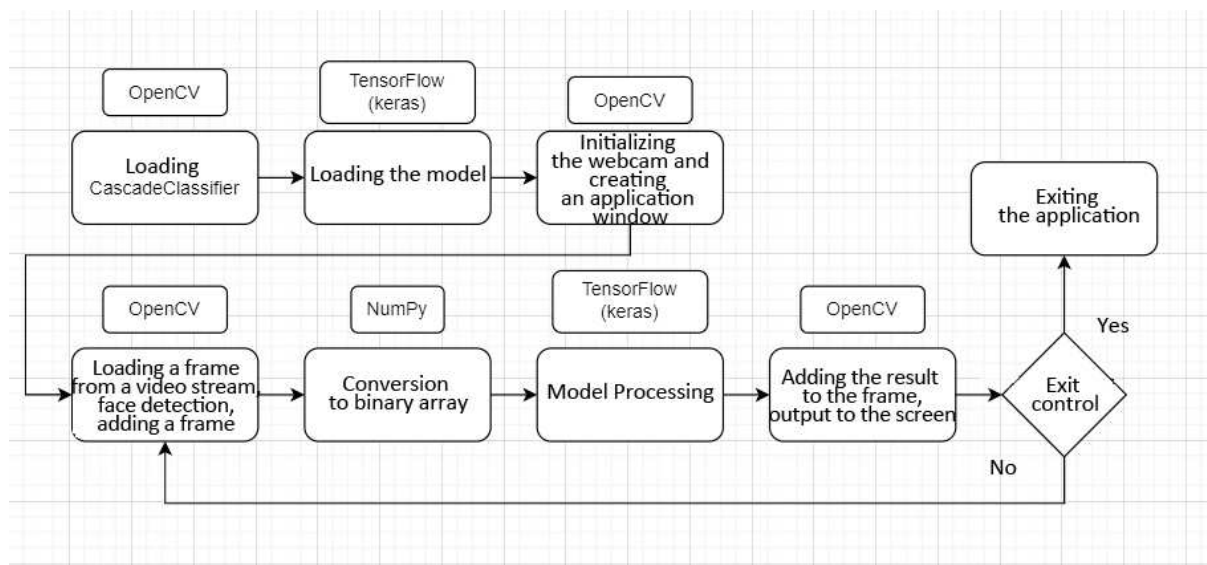


Fig. 4. Interactions of components of the identification subsystem

Figure 5 shows a diagram of the work of the emotion recognition subsystem presented. When calling the *update* method, first the active frame is taken through the *get_frame* method, then the Viola – Jones method determines the boundaries of the face on it, and the frame is transformed for processing and processed by a convolutional neural network. To pass this stage, libraries such as *cv2*, *karas*, *numpy*, and *tensorflow* are used. Next, there is a check for the presence of a frame and for the presence of a tkinter window, in the absence of the program, the program is closed, otherwise, the frames and instance variables are updated, *tkinter* widgets are updated, *matplotlib* graphs are saved and the results of the work are saved to a file using the *threading* library. A separate stream is created, to which the method for saving data to a file and its arguments are passed. At the very end, the *update* method is called with a delay.

When creating an instance of the *VideoCapture* class, a connection to the video stream is made via *cv2*. Also, when calling the *get_frame* method, the active frame is updated using *cv2*.

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.

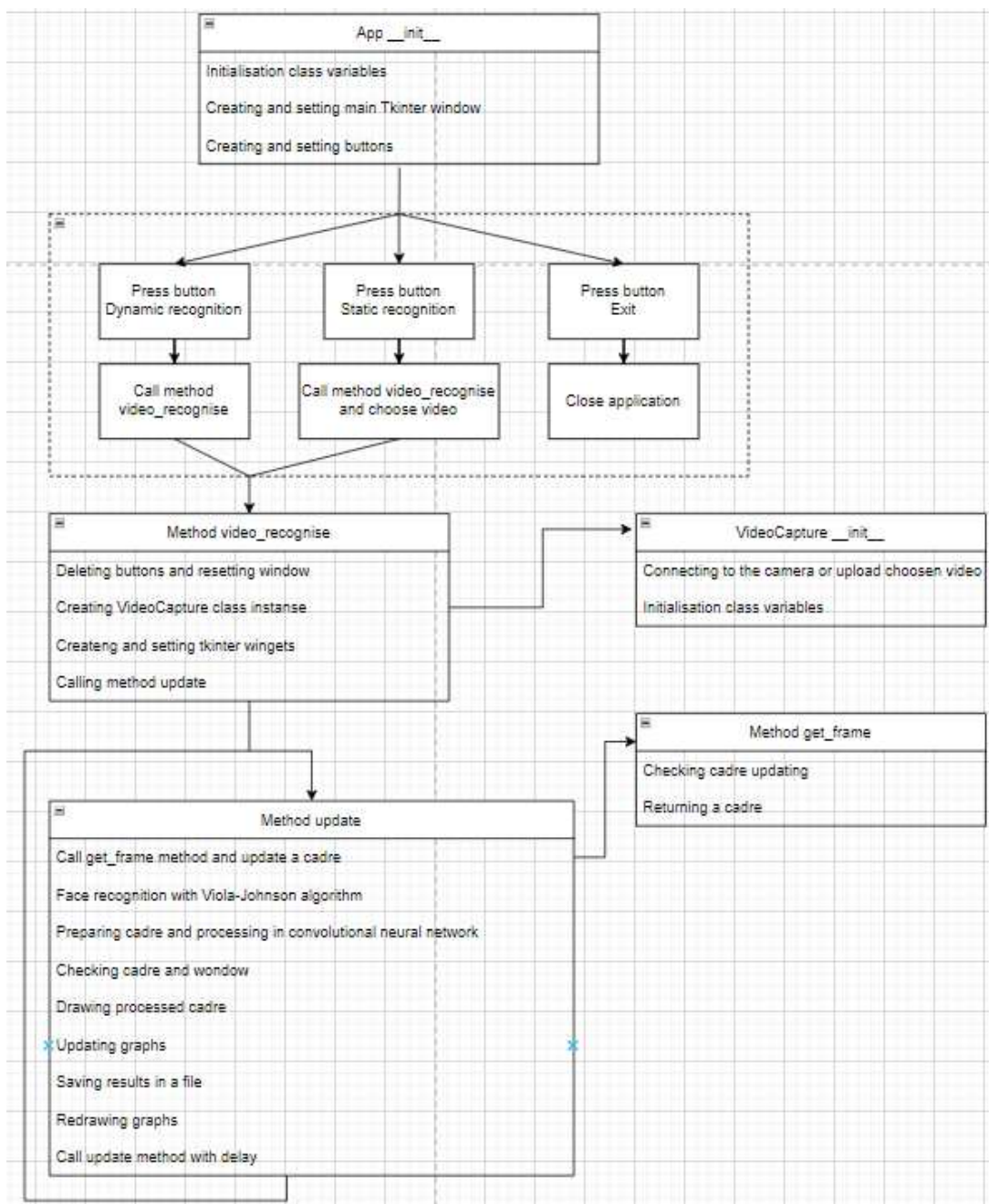


Fig. 5. Application at work

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 (picture 36);
- the *count* variable contains the number of frames that will be displayed on the emotion graph for count frames (Fig. 6);

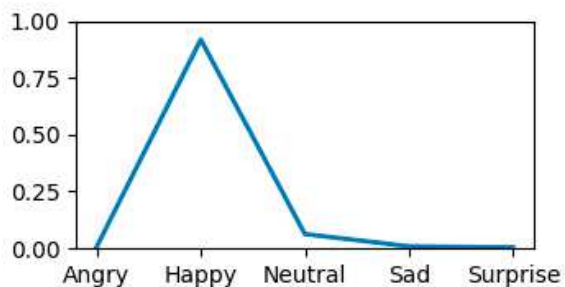


Fig. 6. Active cadre diagram

- 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 (Fig. 7);

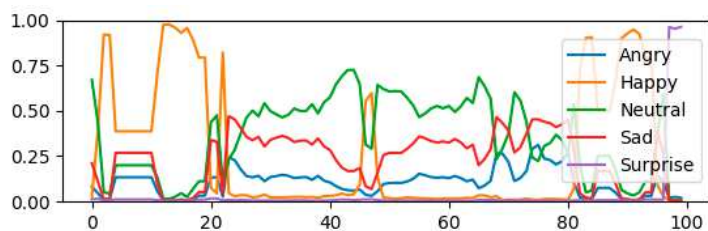


Fig. 7. Diagram for “count” cadres (color online)

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

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

Also in this class, we will create the necessary methods. The `__init__` constructor is used to initialize the main variables of the class instance, as well as to create a selection menu with three buttons: “Dynamic recognition”, “Static recognition”, and “Exit”.

When `__init__` calling, you must pass the following parameters:

- the “window” parameter is a *Tk* class from the *Tkinter* library that creates the main working window;

- the *window_name* parameter is a string variable to create a name in the window’s title bar;

- the “*window_logo_path*” parameter is a string variable that specifies the path to the logo in the program header;

- the *fps* parameter is a limiting parameter that indicates the desired number of frames per second, but due to the possible limitation of the device’s power, it only allows to reduce the load;

- the *video_source* parameter is a parameter with a default value that determines the program’s operation mode. If it is equal to zero, then the definition is based on the video stream from the webcam, and, if not, then get the path and load the video.

Next, the method creates and configures the window, its size, title, protocol, background, and logo. This is followed by the creation of three buttons and their configuration:

- the “Exit” button closes an application after pressing;

- the “Dynamic recognition” button calls *video_recognise* method in web camera mode after pressing;

- the “Static recognition” button calls *video_recognise* method in static video mode after pressing, the user should choose a video for recognizing emotions in the video;

The *video_recognise* method is used to initialize the main graphic widgets, as well as to call the update method. The input is one parameter *video_source*, in which the mode of the program is transmitted. Next, the buttons that were created in the constructor are deleted and the main window is reconfigured. After that, an instance of the *VideoCapture* class is created and the main widgets are created (Fig. 8).

3. Approbation of the system

To implement the project, the *Python* programming language and its libraries were used: *NumPy*, *TensorFlow*, and *OpenCV*. When you run the script, the interface is called, which displays the image from the webcam. When determining the face, an encircling square with an inscription identifying the person’s identity is displayed.

To test the identification subsystem, the trainees were entered into the “*KnownFaces*” database, as shown in Fig. 9.

The program successfully saves the results of the work in separate files for each run. Automatic search, use, and operation of the webcam are carried out (Fig. 10).

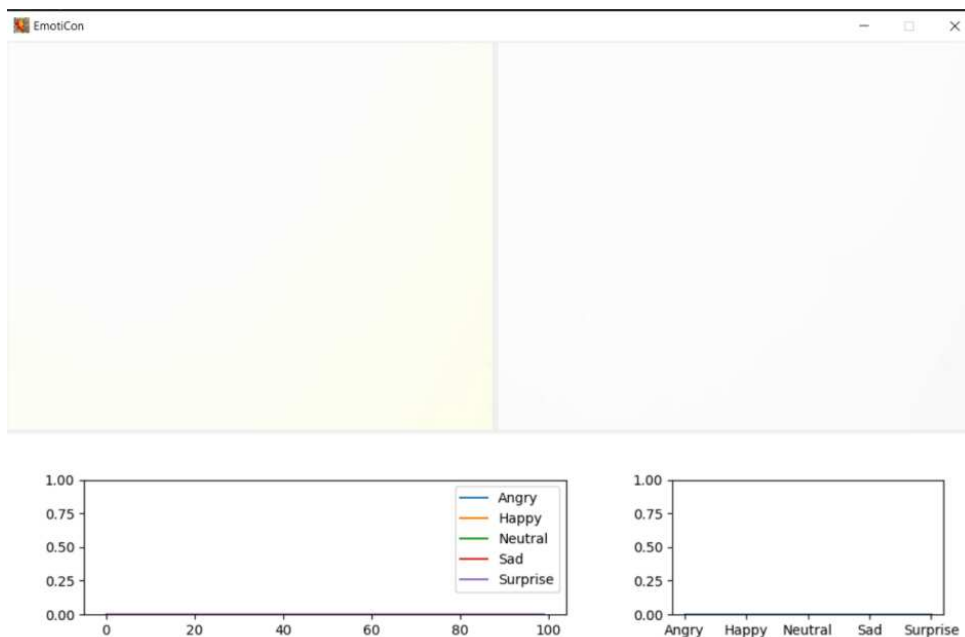


Fig. 8. Main widgets and frames

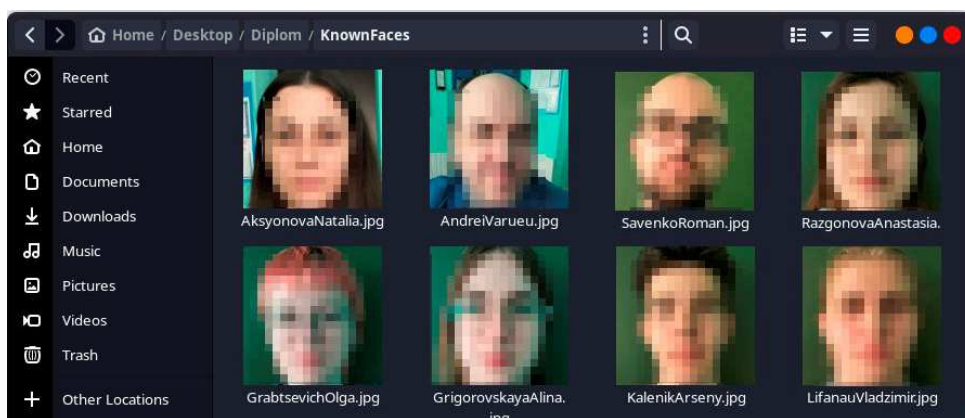


Fig. 9. Adding images to the database



Fig. 10. Testing of the identification subsystem

Testing of the accuracy of the model of the subsystem of recognition of the cognitive behavior of the observed was carried out to determine each emotion. In this test, the ability of the model after training to determine the emotions embedded in the algorithm in different light conditions and from different angles of the face location is tested (Fig. 11–15).

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.

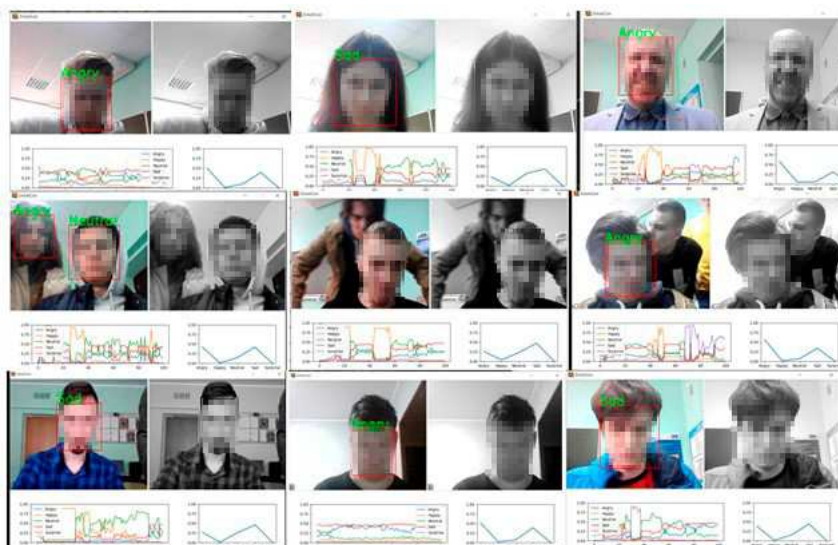


Fig. 11. Testing the emotion "Angry"



Fig. 12. Testing the emotion "Happy"

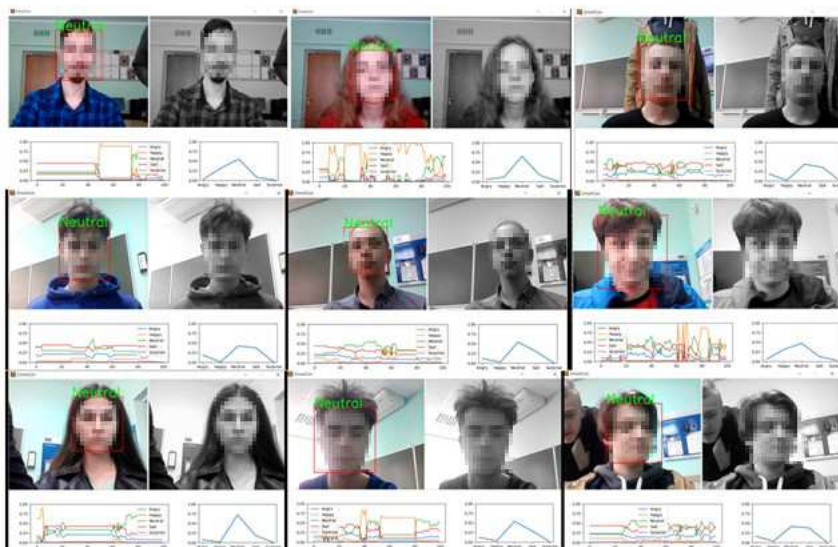


Fig. 13. Testing the emotion "Neutral"



Fig. 14. Testing the emotion “Sad”

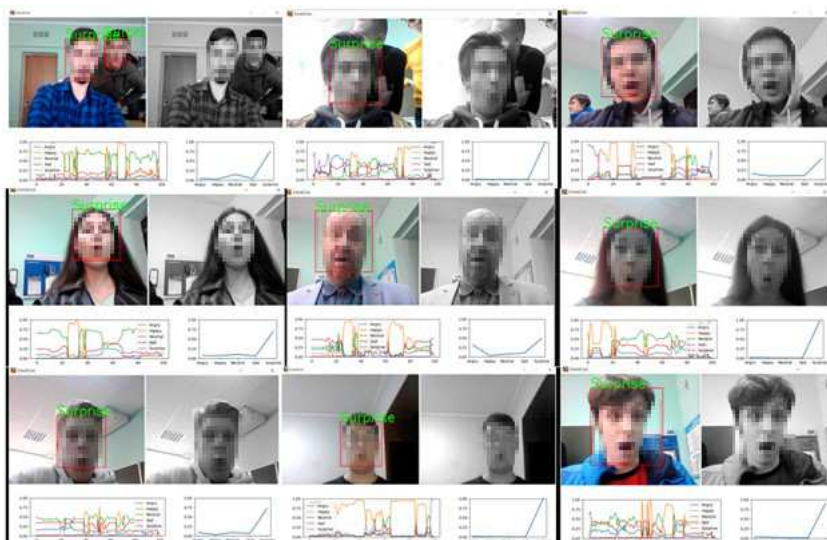


Fig. 15. Testing the emotion “Surprise”

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 both sexes.

Conclusions

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 in the 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 and register the time of arrival and departure automatically. The subsystem for determining the cognitive behavior of the observed requires further refinement.

References

1. Demidenko O. M., Aksionova N. A. Development of a machine vision system for image recognition of design estimates. *Nonlinear Phenomena in Complex Systems*, 2022, vol. 25, iss. 2, pp. 159–167. <https://doi.org/10.33581/1561-4085-2022-25-2-159-167>
2. Badrinarayanan V., Kendall A., Cipolla R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, vol. 39, iss. 12, pp. 2481–2495. <https://doi.org/10.1109/TPAMI.2016.2644615>
3. Viola P., Jones M. Rapid object detection using a boosted cascade of simple features. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2001, vol. 1, pp. 511–518. <https://doi.org/10.1109/CVPR.2001.990517>
4. Shapiro L., Stockman G. *Computer Vision*. London, Pearson, 2006. 752 p.
5. Aksionova N. A., Demidenko O. M., Voruev A. V. Implementation of a system for determining students' emotions by their facial expressions. *Proceedings of Francisk Skorina Gomel State University. Natural Sciences*, 2022, iss. 3 (132), pp. 82–87 (in Russian).

Поступила в редакцию / Received 15.12.2022

Принята к публикации / Accepted 19.02.2023

Опубликована / Published 31.05.2024