Emotion Recognition System with Facial Expression System for a Vehicle to Make the Journey Harmless

Haju Mohamed Mohamed Naleer and Narmadi Wathsala Dissanayake

Faculty of Applied Sciences,
South Eastern University of Sri Lanka,
Sammanthurai, Sri Lanka
hmmnaleer@seu.ac.lk

Abstract. Image processing is commonly used to recognize objects. It can be used to detect human facial expressions and emotions by observing facial features. Capturing emotions depends on many tools and conditions such as brightness, camera, changing rate of facial features etc. Emotion recognition is software that allows program to "read" the emotions on the human face using advanced image processing. In order to understand not only what a person's face or image looks like, but also how it looks. Emotion recognition has been applied in many fields such as Medical, Security, and Business etc. There are many difficulties in developing a good emotion recognition system for the human face in real time. Since most of the time facial features of expression and the style of showing emotion to the outside world is different from person to person. Therefore, it is very difficult to build an accurate system for real time emotion recognition. This project is to detect human facial expressions to predict the current emotional state. The system specially focused on reducing fatal road accidents due to drivers' state of emotion. Initially it is built to recognize the human emotion through facial expressions and then evaluated to detect drowsiness using facial landmarks to ensure the safety of the driver. Training has been done with Kaggle dataset for seven emotional states (Neutral, Happy, Angry, Sad, Scared, Surprised and Disgust) called universal emotions. In order to predict drowsiness, it uses specific 12 points on face (6 points on each eye) in shape predictor 68 face landmarks. Evaluated system has given 66% accuracy in testing and drowsiness alert also showed a very good success rate.

Keywords: Highways Traffic Surveillance System, IP camera, OpenCV

1 Introduction

Companies are experimenting with integrating modern algorithms with image processing techniques that have emerged over the last 10 years and also the probabilities of the corresponding emotions. The change of feelings from time to time directly affects the performing activity. Therefore, human emotions are very important to recognize and understand the situation of the respective person. Recognition of human emotion can be a very easy and normal task for another human being but not for an AI. Nowadays this topic is very popular among the field of AI and most of the software companies tend to develop new and highly accurate algorithms for the purpose of detecting human emotions. Apart from that there are many more fields that already use this technology to reduce human effort and problems. Emotion detection system for a driver is one of the applications among those fields.

It is obvious that risky behaviour of drivers significantly increases the probability of an accident. It has been claimed that driving-related errors cause around 95% of fatal accidents.

© Faculty of Management and Commerce
South Eastern University of Sri Lanka (SEUSL)
University Park, Oluvil 32360. Sri Lanka
Received: 13th, 2020
Revised and Accepted: 21st Jul, 2020

Most of the time we happen to make some risky behaviours including speeding, getting distracted (e.g. using mobile phones, looking around, chatting) while driving with or without noticing. This project introduces an emotion detection system to recognize the present state of the driver, since it can represent the feelings of a human being in different situations. In the system, it has combined the science of psychology, human expressions and AI to recognize different emotions of an individual's face automatically. Naturally the facial expressions reveal the inner emotions accurately than words or other actions. While communicating, humans are able to produce thousands of facial expressions that are different from one another in intensity, complexity and also in the meaning. The system classifies seven emotions (Angry, Scared, Sad, Disgust, Happy, Surprised and Natural) with their probability values in order to clarify if the emotion is risky or not. If it is risky the system may send personalized alerts (Ex: asking him to stop for a coffee break, change music or temperature) to reduce the tendency of road accidents while increasing the safety.

Normally human emotions change frequently, maybe in minutes. Most of the time the changing emotions may not be harmful for the journey than the ones which last for a considerable period of time. Other than that, making alerts for each and every emotion which changes naturally will be a sort of disturbance for the motorist while it is not even useful because the current state will change before the alert.

Simultaneously the system detects the eye motions using specific 12 points of facial landmarks (6 points for each eye) and calculate the eyes' average blinking ratio in order to estimate the drowsiness of the driver. If the system is satisfied with the given arguments the drowsiness alert will turn on saying "wakeup, please". The accuracy and performance of the proposed system can also depend or vary with the following:

- Brightness of the surroundings.
- The quality of the camera used for video capturing.
- Person dependencies. (Style of showing emotions through facial features)
- Adverse conditions

The main goal in this project is to make the journey harmless or safe and reduce the life risks. At the same time, it can save drivers from getting charged for road violations that occur due to unexpected mistakes. Also, the system can make the driver confident in his journey while reducing the road accidents. Through this system, the driver is directed to take immediate actions for his/her present emotional state and helps to have self-control against the emotions. This kind of automated systems do have more efficiency rather than using a human inspection system which includes continuous human observations.

This paper consists literature review which discussed the referenced documents/articles or other resources. Next, we have specified the methodology that we went through. Results of this research, conclusion and future works will take a part at the end of the paper respectively.

2 Literature review

[1] fulfil the recognition of facial action units like the subtle change of facial expressions, and emotion-specified expressions. Face detection, facial component extraction, tracking and representation, and facial expression recognition includes in the system to analyse facial expression. Canny Edge Detector, the optimum facial feature extraction algorithm, is applied to localized face images, and a hierarchical clustering-based scheme reinforces the search region of extracted highly textured facial clusters. The results of the system show that it can

be well applied to real-time facial expression and emotion categorization tasks. In [2] it discussed a software which detects and recognizes faces as well as tells a lot more about that person which could be used to get feedback from customers or to know if a person needs motivation has presented. Being an affordable and efficient product is the objective of the project. Python, AI, image processing technology are primarily used to develop the system. In order to avoid any accident or miss-happening mainly eye blinking is detecting [2].

For automatic real-time facial expression recognition is undertaken along with their benefits and flaws, to comparative study of the different approaches initiated for automatic real-time facial expression recognition in [3].

[4] analyses the strengths and the limitations of systems based only on facial expressions or acoustic information. Two approaches are discussing, decision level and feature level integration. Sadness, anger, happiness, and neutral state emotions are classified using a database recorded from an actress. The results say that the system based on facial expression gave better performance than the system based on just acoustic information. [5] Presents a multimodal fuzzy inference system for emotion detection. Visual, acoustic and context relevant features extracts and merges by the system. Based on an invariant representation of facial expressions, the system proposes a method that adapts to the morphology of the subject. Method relies on 8 key expressions and each image of a video sequence is defined by its relative position to these expressions. Expression recognition is performed with a basic intensity-area detector. Valence, arousal, power and expectancy are the 4 dimensions that emotion is described. Results show that the duration of high intensity smile is an expression that is meaningful for continuous valence detection and can also be used to improve arousal detection.

A technique [6] that was introduced by two researchers Viola and Jones to accurately and rapidly detect faces within an image and it can be adapted to accurately detect facial features. False positives are eliminated by regionalizing the detection area and the speed of detection is increased due to the reduction of the area examined. Based on still images of the face paper discuss a framework for the classification of emotional states in [7]. Parameters from the AAM are used as features for a classification scheme that is able to successfully identify faces related to the six universal emotions. The effectiveness of AAMs in capturing the important facial structure for expression identification and also help suggest a framework for future development.

[8] Investigates various emotion recognition techniques from the facial expression of human subjects. Active shape models (ASMs), and translated 49 scalar features used to describe human facial expressions, in that case a number of characteristic points are extracted from input face images. Support vector machine (SVM) and multi-layer perceptron (MLP) statistical pattern recognizers, are used to identify various emotions from the feature vectors. 5-fold cross-validation is carried out, with varying numbers of emotions from 3 to 6 to analyse the performances of the various pattern recognizers. SVM is the most stable and best in terms of emotion classification rates according to the results.

[9] Proposes a method for generating a subject-specific emotional feature space to express the correspondence between the changes in facial expression patterns and the degree of emotions. Self-organizing maps and counter propagation networks used in generating the feature space. Investigated the training data input method and the number of dimensions of the CPN mapping space. The results clearly show that the input ratio of the training data should be constant for every emotion category and the number of dimensions of the CPN mapping space should be extended to effectively express a level of detailed emotion.

In [10] it has been used deep learning techniques in order to approach learning several specialist models. Explore multiple methods for the combination of cues into one common classier. This achieves a considerably greater accuracy than predictions from the strongest single-modality classier. The method was the winning submission in the 2013 EmotiW challenge and achieved a test set accuracy of 47.67% on the 2014 dataset. Comparison of two implementations of emotion recognition in faces are compared in [11]. Performance between OpenCV algorithm for emotion recognition and an implementation of Emotion cognitive service are testing with the use of two Android mobile applications and one thousand tests were carried out per experiment. Results gives that OpenCV implementation got a better performance than the Cognitive services application. Performance can be improved by increasing the sample size per emotion during the training step in both cases.

[12] Presented some novel methods for facial emotion recognition in facial image sequences. System is based on Active Appearance Models (AAM) with Ekman's Facial Action Coding System (FACS). AAM and DBN, can achieve a higher recognition performance level by combining them, when compared with other facial expression recognition methods. Shows more than 90% accuracy. The work [13] is detecting electroencephalogram (EEG) signals and facial expressions continuously as specified in the paper through video. Valence levels are provided continuously by annotators using frontal facial videos of participants. The process used Power spectral features from EEG signals as well as facial fiducial points to detect valence levels for each frame continuously. Performance of the model has been verified for emotional highlight detection using emotion recognition from EEG signals and the final results are presented.

[14] Presents the idea of improving the recognition accuracy of one or more of the six prototypic expressions namely happiness, surprise, fear, disgust, sadness and anger. A motion gradient based optical flow for muscle movement is computed between frames of a given video sequence to do so. Generate feature vectors using this computed optical flow and clustering is done with tree generated rule base for recognition of expressions. [15] Proposed an emotions composition model based on the cognitive analysis of independent user's affective facial recognition to dope out what are the user's new affective facial expressions. To show and verify independent user's affective expressions, it has applied the principal component, cluster and discriminant analysis.

3 Methodology

3.1 Architecture

The process of classifying facial expression has to deal with some variables of the human face like colour, texture, posture, expression, orientation and so on. The recognition is done based on the muscle movements and eyes, nose and lips have to be detected as it needs to focus on nonverbal signals to understand the exact emotion by comparing the current state with trained data set and then it can be classified under specified categories. The system shown in Fig.1. In the system, it has combined the science of psychology, human expressions and AI to recognize different emotions on an individual's face automatically. The system classifies seven emotions (Angry, Scared, Sad, Disgust, Happy, Surprised, Natural) called universal emotions, in order to clarify if the emotion is risky or not. If it is risky the system may send personalized alerts.

Other than that, there is a process to detect the drowsiness of the driver to pretend from the accident rate and it observes the specific points around the eye from facial landmarks to detect whether the eyes are opened, closed or blinking. If eyes are closed for a considerable time as specified the system should alert the driver to stay awake. System will consist of multiple stages like image capturing, image processing, expression detection, emotion recognition, eye blinking detection, analysis of the risk probability, selection and passing the appropriate voice alert etc.

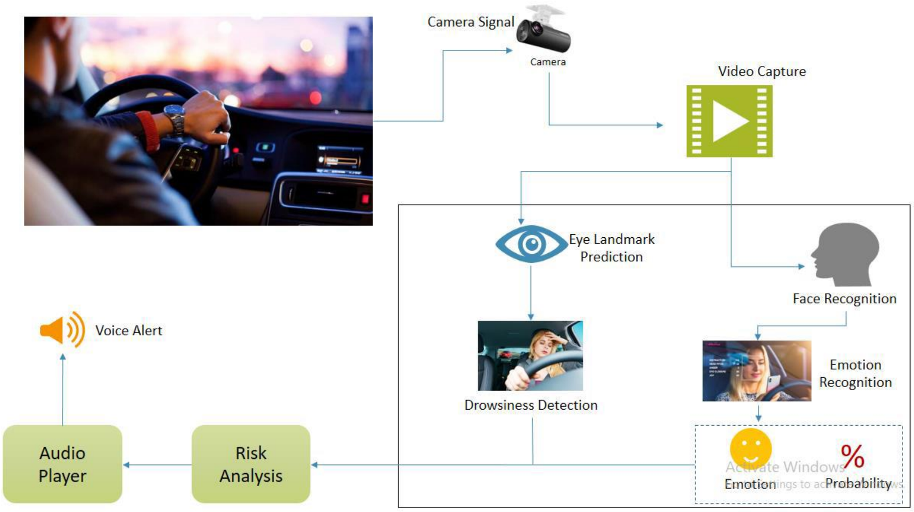

Fig. 1. Design of the Proposed System (Flow Chart)

3.2 Emotion recognition

The main part of the system is running in real time using a camera to take videos. After capturing the image using a camera, the system recognizes the face with image processing and expression detection process will proceed. Classification of the emotion is done according to the detected expressions based on the trained dataset. Seven emotions are detected by using facial features and calculates the probability of the current state. The one that takes the highest value is the emotion. Emotions which are mostly affected for a risk are specially considered to make an alert if it has a tendency to create a problem as defined by the system. The colour and gray-scale formats of frames, face detection and emotion recognition were shown in Fig. 2, Fig.3, and Fig. 4 respectively.

Fig. 2. Colour and Gray-scale formats of frames

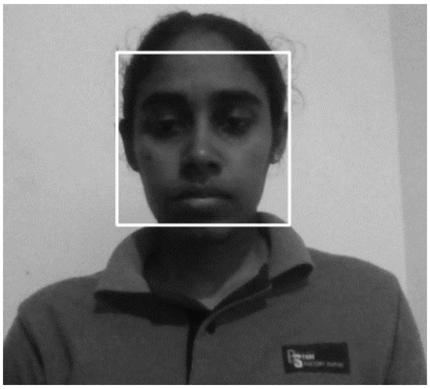

Fig. 3. Face Detection

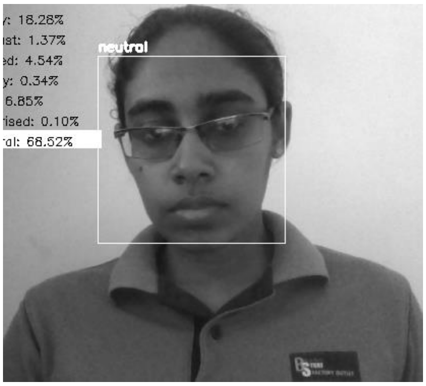

Fig. 4. Emotion Recognition.

3.3 Drowsiness Detection

The drowsiness detection process also carried out simultaneously by observing the eye landmarks to detect the specific 6 points around the eye and calculates the vertical and horizontal values of both eyes. Using those calculated lengths average ratio can be compute and based on that alert will operate when necessary.

For the purpose of calculating horizontal and vertical lengths of the eye for detection of drowsiness, we are using 12 points of landmarks (6 points for each) around the eyes out of 68 facial landmarks (see Fig. 5 and Fig. 7). By using those landmark points vertical and horizontal lengths can be detected. The shape prediction is shown in Fig. 6.

Fig. 5. Shape Predictor 68 Face Landmarks

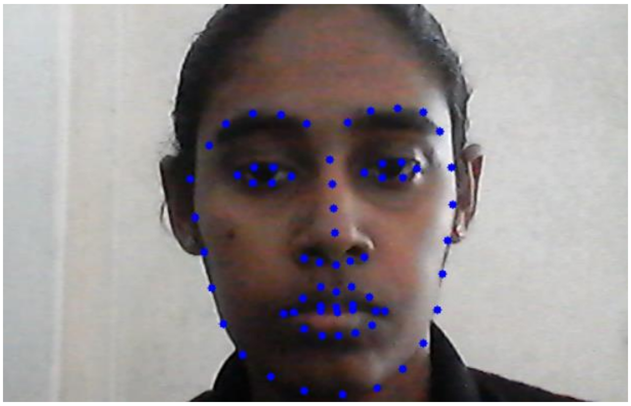

Fig 6. Shape Prediction

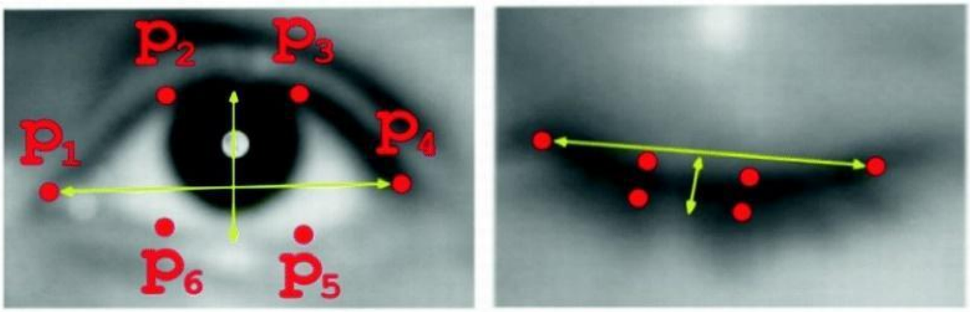

Fig. 7. Landmarks around the Eye and Horizontal and Vertical Lines (Detect the Length and Calculate the Ratio)

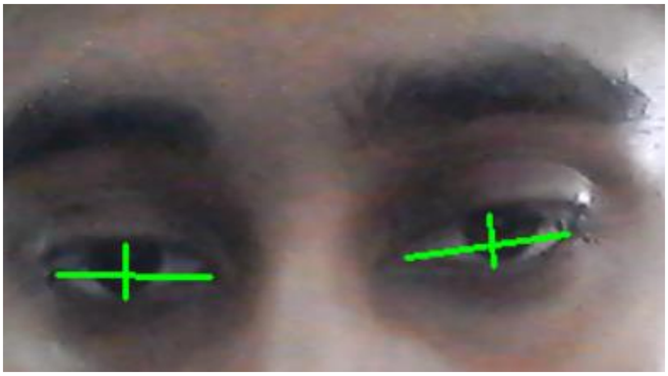

Fig. 8. Drawing Horizontal and Vertical Lines Around Eye

- Top mid-point = (P1 + P2)/2
- Bottom mid-point = (P6 + P5)/2
- Vertical length = length from top mid-point to bottom midpoint (Fig. 8.)
- Horizontal length = length from point P1 to point P4 (Fig. 8.)

(Above points are referred from the Fig. 7)

On the purpose of reducing the errors, the ratio has to be calculated with ratio values of both eyes as average ratio value. If the average ratio is increased specific predefined value that given by the system, blinking has to be detected. Whenever the duration of the blinking is extending the specified duration, it is going to be detected as drowsiness and should turn on the alarm or send an alert to the user.

4 Results

4.1 Results in emotion recognition

System is recognizing five emotions (Happy, Neutral, Sad, Angry, and Scared) accurately and other two are not much accurate.

Table 1: Actual Emotion vs. Recognized Emotion with Analysed Probabilities

Probability of Emotions (100%)							Recognized Emotion	Actual Emotion
Neutral	Happy	Angry	Sad	Surprise	Disgust	Scared		
Face 01								
68.52	0.34	18.28	6.85	0.10	1.37	4.54	Neutral	Neutral
7.48	**85.19**	6.42	0.18	0.13	0.03	0.56	Happy	Happy
22.62	2.05	**47.65**	14.91	0.31	0.76	11.70	*Angry*	Disgust
18.11	1.23	8.18	3.58	19.05	0.01	**49.84**	Scared	Scared
4.47	0.83	**86.90**	0.52	1.60	0.11	5.57	Angry	Angry
15.42	4.87	22.87	**26.46**	4.02	0.25	26.11	Sad	Sad
18.89	15.46	12.42	17.86	8.44	0.80	**26.12**	*Scared*	Surprise
Face 02								
30.42	13.93	16.39	24.73	1.05	0.03	13.46	Neutral	Neutral
16.03	**30.85**	28.82	5.99	2.79	0.09	15.43	Happy	Happy
19.13	**33.96**	8.99	24.57	0.87	0.06	12.43	*Happy*	Disgust
18.82	8.43	25.21	8.07	8.30	0.10	**31.07**	Scared	Scared
0.59	18.46	**65.45**	0.48	7.27	0.20	7.56	Angry	Angry
18.82	2.69	15.41	**48.35**	0.40	0.06	14.27	Sad	Sad
15.72	7.32	3.93	1.93	**56.94**	0.00	14.17	Surprise	Surprise

4.2 Results in Drowsiness Detection

System designed to detect eye blinking and closing states. If both eyes are closed for a period of time, system gives voice alert saying “wake up please”. Only when the user continues closing his/her eyes system is giving the voice alert saying “wake up please. You are feeling dizzy, let’s stop for a coffee”. System is able to detect linking accurately (over 90%) even when wearing spectacles.

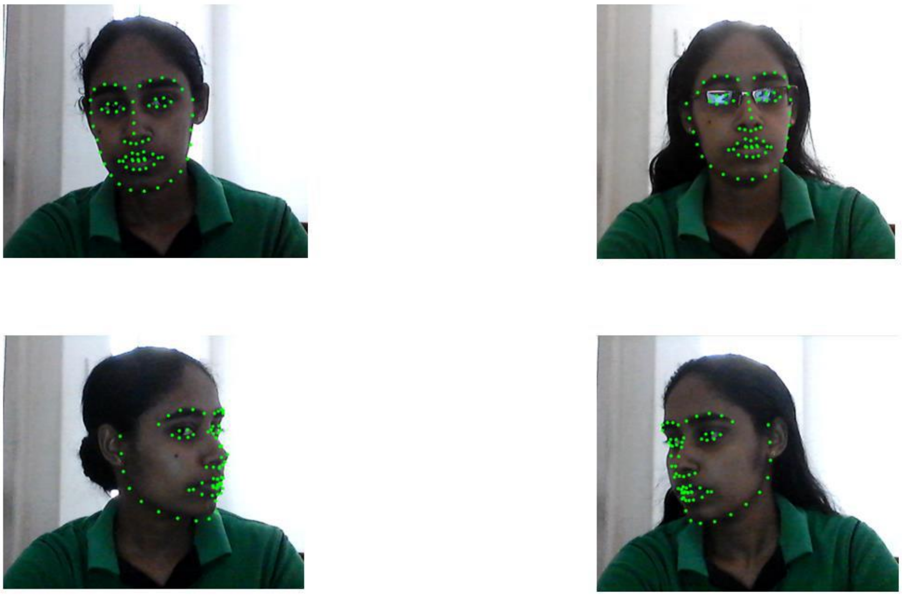

Fig. 9. Landmark Detection With and Without Spectacles and In Angles

5 Conclusion

System is built to be a solution for a common problem of fatal road accidents. It has developed using python programming language and OpenCV is used for real time video capturing. Camera and an audio player or speaker should be used to implement the system as hardware objects in order to get input for and play output from the system. In designing phase, Tensorflow, Keras and also Haar cascade classifiers to specify the emotions. Kaggle emotion detection dataset that consists of 48×48-pixel grayscale images of faces were used to train the system. Drowsiness detection is done by using shape predictor 68 face landmarks.

By the end of this project we were able to implement a system to reduce the risk of happening road accidents due to human (motorists) emotions and drowsiness. It does not require much hardware equipment other than a camera to capture live video.

It is better to have a camera with good quality to get videos with high resolution in order to detect the emotions with higher accuracy. For that case the requirement of the capacity and the computational power will be increased.

Environmental dependencies such as light, weather etc. will also can be an effect to the system for reducing its performance. When there is low light surrounding, the camera may not be capable of capturing the real time video and it may cause the system to stop working. This situation widely will take place at night time when the system is mostly required since motorists usually tend to feel sleepy while driving at night.

Other than the above mentioned problems, there are some effects for the system to be implemented accurately with respect to the user. Nowadays most of the drivers are used to having sunglasses or some other vision protection items while driving. This can cause a big issue to the work system in proper manner since it is a huge barrier for reading specific points on the face to detect eye landmarks and expressions.

Another dependency for the accuracy of the system can classify as the style of the face. That means, the natural human faces are different from one another. Some are different from shape when some are from colour. There are some people who have closed eyelids even when they open their eyes (Ex: Chinese people). With these types of users, the system is calculating the average ratio using eye landmarks and may bring an output as the eye is closed if the defined conditions of the system are satisfied.

Face covering is another difficulty for detecting the human facial expressions since it covers all the required points on face for the detection process. Hence the recognition of emotions and drowsiness is problematic.

There are some other factors which can affect the system performance such as makeup, hairstyles that cover the eyes or part of the face, wearing caps and hats or scarves etc.

6 Future Works

With the successful detection of facial features, the next goal is to increase the accuracy of the system with advanced feature extraction methods by using a large number of landmarks. Also, it can focus on recognizing emotion states more delicately.

And the project is planned to develop in a way that can work on low light conditions like night time. In brief, the above-mentioned problems are going to be considered and make some sort of solutions that can overcome those problems.

References

[1] Ghahari A, Fatmehsari YR, Zoroofi RA. A novel clustering-based feature extraction method for an automatic facial expression analysis system. In2009 Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing 2009 Sep 12 (pp. 1314-1317). IEEE.

[2] Uppal A, Tyagi S, Kumar R, Sharma S. Emotion recognition and drowsiness detection using Python. In2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence) 2019 Jan 10 (pp. 464-469). IEEE.

[3] De A, Saha A. A comparative study on different approaches of real time human emotion recognition based on facial expression detection. In2015 International Conference on Advances in Computer Engineering and Applications 2015 Mar 19 (pp. 483-487). IEEE.

[4] Busso C, Deng Z, Yildirim S, Bulut M, Lee CM, Kazemzadeh A, Lee S, Neumann U, Narayanan S. Analysis of emotion recognition using facial expressions, speech and multimodal information. InProceedings of the 6th international conference on Multimodal interfaces 2004 Oct 13 (pp. 205-211).

[5] Soladié C, Salam H, Pelachaud C, Stoiber N, Séguier R. A multimodal fuzzy inference system using a continuous facial expression representation for emotion detection. InProceedings of the 14th ACM international conference on Multimodal interaction 2012 Oct 22 (pp. 493-500).

[6] Wilson PI, Fernandez J. Facial feature detection using Haar classifiers. Journal of Computing Sciences in Colleges. 2006 Apr 1;21(4):127-33.

[7] Ratliff MS, Patterson E. Emotion recognition using facial expressions with active appearance models. InProc. of HRI 2008 Mar 17.

[8] Jang GJ, Park JS, Jo A, Kim JH. Facial emotion recognition using active shape models and statistical pattern recognizers. In2014 Ninth International Conference on Broadband and Wireless Computing, Communication and Applications 2014 Nov 8 (pp. 514-517). IEEE.

[9] Ishii M, Shimodate T, Kageyama Y, Takahashi T, Nishida M. Generation of emotional feature space for facial expression recognition using self-mapping. In2012 Proceedings of SICE Annual Conference (SICE) 2012 Aug 20 (pp. 1004-1009). IEEE.

[10] Kahou SE, Bouthillier X, Lamblin P, Gulcehre C, Michalski V, Konda K, Jean S, Froumenty P, Dauphin Y, Boulanger-Lewandowski N, Ferrari RC. Emonets: Multimodal deep learning approaches for emotion recognition in video. Journal on Multimodal User Interfaces. 2016 Jun 1;10(2):99-111.
[11] Prieto LA, Komínkova-Oplatková Z. A performance comparison of two emotion-recognition implementations using OpenCV and Cognitive Services API. InMATEC Web of Conferences 2017 (Vol. 125, p. 02067). EDP Sciences.
[12] Ko KE, Sim KB. Development of a Facial Emotion Recognition Method based on combining AAM with DBN. In2010 International Conference on Cyberworlds 2010 Oct 20 (pp. 87-91). IEEE.
[13] Soleymani M, Asghari-Esfeden S, Pantic M, Fu Y. Continuous emotion detection using EEG signals and facial expressions. In2014 IEEE International Conference on Multimedia and Expo (ICME) 2014 Jul 14 (pp. 1-6). IEEE.
[14] Ahmed W, Mitra S, Chanda K, Mazumdar D. Assisting the autistic with improved facial expression recognition from mixed expressions. In2013 Fourth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG) 2013 Dec 18 (pp. 1-4). IEEE.
[15] Chao X, Zhiyong F. Facial expression recognition and synthesis on affective emotions composition. In2008 International Seminar on Future BioMedical Information Engineering 2008 Dec 18 (pp. 144-147). IEEE.