Detection of Emotional States Through the Facial Expressions of Drivers Embedded in a Portable System Dedicated to Vehicles

Eduard Zadobrischi[1,2], Lucian-Mihai Cosovanu[1,3], Mihai Negru[2] and Mihai Dimian[1,3]
[1]Department of Computers, Electronics and Automation, Stefan cel Mare University of Suceava, Romania
[2]Technical University of Cluj-Napoca, Cluj-Napoca, Romania
[3]Integrated Center for research, development and innovation in Advanced Materials, Nanotechnologies, and Distributed Systems for fabrication and control, Stefan cel Mare University of Suceava, Suceava, Romania
eduard.zadobrischi@usm.ro

Abstract— **The density of cars and the technological advance in the last decade of this century have increased exponentially, reaching a point that creates various problems. The main cause of deaths due to car accidents has a lesser-known cause and that's coming from the driving area under certain conditions and under certain emotional states that negatively or positively affect the driving style. In this article, we present a module implemented in a multifunctional device dedicated to cars that aim to detect the emotional states of drivers. Thus, depending on these conditions, the application specifies the driver's status by identifying certain features that can later prevent him from driving the car until the emotional state is appropriate and does not pose a danger to their own lives and to other road users or pedestrians.**

Keywords— face detection, emotion detection, vehicle safety applications, infrastructure-to-vehicle communications, safety driving.

I. INTRODUCTION

In terms of statistics, we have about 50 million people injured annually in road accidents, which are a major problem in today's society, producing the loss of human lives and economic damages. Current technological development and advancement have always offered solutions and approaches in both safety systems and smart cars. Most developed systems reduce some of the effects but do not act on the root cause around which the whole situation revolves. The driver is the main subject as a given unpredictable human factor and often influenced by the emotional state in the process of driving the car. This emotional characteristic positively or negatively affects the driving style and automatically the effects that a driver with a certain emotional state can produce can sometimes be tragic [1]. Thus, human features and characteristics but also emotional states are different, appearance, movements, facial expressions, facial grimaces, facial deformities but also the tightening and hiding of certain state contracts an ideal direction in research and analysis.

Facial expressions are some of the strongest and most natural signs by which human beings transmit their emotional states. Even in the field of computer vision and autonomous learning, different expression recognition facial (ERF) systems are used to explore coded information and facial expressions. Since the twentieth century, Ekman and Friesen [2] have defined six basic emotions on the foundations of the study of intercultural, which clearly indicates that people perceive certain basic emotions but in different ways, depending on the culture they come from. We have defined six basic emotions on the foundations. These expressions differ from one culture to another but the most important ones found in most cultures have prototypical characteristics as anger, disgust, fear, happiness, sadness, and surprise with small derivations depending on facial features. We can say that advanced research on neuroscience and psychology claims that there are six basic emotions that are found in different cultures but are not universal. Being an extremely volatile area through the unpredictable analysis of the human factor there is an extremely clear direction in the development of applications dedicated to traffic safety, pedestrians, and drivers. Starting with 2013, massive studies and procedures for collecting raw emotions and testing them in challenging real-world scenarios began to highlight the degree of influence in a practical case and what could be the damage caused by such an event. As the first step in this paper, we have created an application and a portable system for the analysis of still images and video captures in which driver's features are identified, highlighting a certain mental state he goes through, being directly influenced by reflecting in personal driving style. The developed module is part of a system dedicated to traffic safety, optical communications using revolutionary and fast technologies such as visible light communication (VLC), vehicle-to-everything (V2X), vehicle-to-vehicle (V2V), vehicle-to-infrastructure (V2I), or dedicated short-range communication (DSRC) [3].

II. DEBATE ON THE EXPERIMENTAL TESTING SETUP

For many people, emotional state condition and impose certain behavioral reactions, these often being conditioned responses by the state we go through. Thus, we often have uncontrolled reactions, and very often they are unfortunate, especially when it comes to driving, an exact activity that can cost our lives, both ourselves and others if we perform negligent movements and totally uncontrolled. We can say that a large part of human behavior is influenced and determined by emotion and not by reason. According to the literature and analysis conducted by Dr. Peter Noel Murray, which describes

This work was supported by a grant of the Romanian Ministry of Research and Innovation, CCCDI - UEFISCDI, project number PN-III-P1-1.2-PCCDI-2017-0917, contract 21 PCCDI/2018.

978-1-6654-0499-0/20/$31.00 ©2020 IEEE

the roles of emotion and behavior of people using magnetic resonance equipment (MRI) whose main role is brain activity and emotion, feelings, deeds. Therefore, emotions force human beings to act and react, if we have states of anger or frustration, anger, they are reflected in our style of expression and the way we drive (see Figure 1) [2].

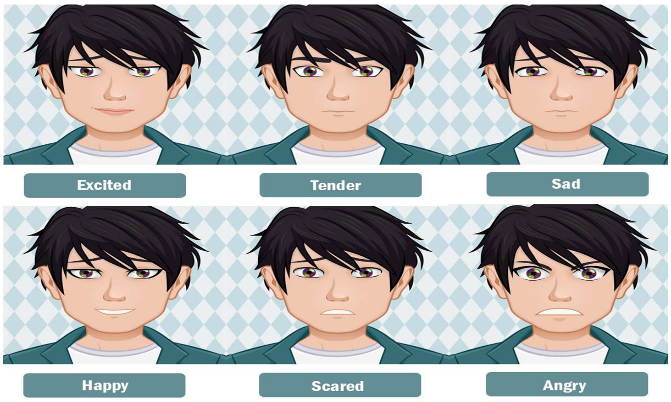

Fig. 1: Facial expressions that can influence the mental state

Thus, human beings are emotional creatures (see Figure 1) and these emotions coordinate our behavior whether we want to or not. A happy person can drive at a pace in which he has control of the car as opposed to an unhappy one who can sometimes have the tendency to exaggerate and go to extreme, thus becoming very dangerous.

A. *Intro – Discussions on the experimental setup*

We used to develop the module and platform programmable commercial components such as Raspberry Pi 3 and 4, Teensy 3.6, and Raspberry Pi W microcontrollers, interconnected using designed components. It is part of a complex platform dedicated to safe driving, augmented reality, inter-vehicular communications. The system of which this module is part is presented in Figure 2, we can see the modularity and flexibility of the system through adaptable center console and most dimensions are adjustable being compatible with any car.

Fig. 2: Portable touchscreen system used for the emotional detection.

For the display and user interface we use a dedicated Raspberry Pi4 touchscreen, and the rest of the components are 3D printed as needed. The developed platform also performs communications such as vehicle-to-vehicle (V2V), vehicle-to-infrastructure (V2I), vehicle-to-everything (V2X) but also visible light communications (VLC). This module can constrain the driver following the facial and emotional analysis, thus prioritizing certain layers depending on the degree of influence on the driving style. Emotional states are scanned and analyzed for each new set of movements detected, thus reducing the states of nervousness or anxiety that can cause road accidents and automatically victims of these events.

B. *Discussion on the detection algorithm for facial expression*

The algorithm for emotional processing and analysis is a new type of algorithm for end-to-end detection, after region-based convolutional neural networks (R-CNN), Fast R-CNN, and Faster R-CNN [4],[5]. It manages to combine classification and detection by regressing the position and the analyzed area of the detection frame in the final layer, turning everything into a regression problem [6].

Thus, by using K-means we lose in terms of accuracy in selecting the starting points and in terms of time because more tests are needed to validate a final and ideal solution. The use of this algorithm comes to our aid to concrete elements that have a similarity between facial movements, some being common and the result can be flawed [7],[8]. In order to be able to have a punctual analysis between the distance detected by the function and the predictability zone compared to the standard defined by Formula (1), where IOU represents that ratio between the intersections between the sets that unify the predicted area with the real area.

$$d = \sum_i \sum_j 1 - IOU(box_i, truth_j) \qquad (1)$$

According to the literature for a proper analysis and detection process we must follow some specific steps in terms of the maximum number of iterations, setting the flight speed for each conjunction $v = 1$; using the K-means algorithm to group the data and to be able to obtain the central m cluster. Then we calculate the degree of individual sharing for each element in the group, and the common function for calculating the distance is Formula (1). Thus, if we have a shorter distance, the value we share is automatically higher. After calculating the degree of sharing for each element we can adjust the degree of fit for each conjunction of the analyzed area using the following formula:

$$F_i = \sum_{i=1}^{m} d_i \;\; i = (1,2,3 \ldots . m) \qquad (2)$$

The next step is that in which the results are ordered in ascending order according to the degree of fit of each element analyzed more precisely individual in the form of the first n individuals ($n<m$); thus a proportional selection operation is obtained and afterwards the cross-selection and the uniform calculation of the variation in the form $P(t)$ uniformly calculating the variation to obtain $P_i(t)$.

Combining n and t permutations for individuals scanned into memory is inserted into a new clustering process $n+t$. Comparing the physical form of individuals and processing groups below the degree of matching we can also implement a filtering or penalizing function $Fmin\ (x_i, x_j)$. Then repeat the previous step and apply the Formula (2) for updating evolutionary algebraic memory *e=e+1* until it reaches the highest number of iterations and tends to the smallest number of elements with the best performance. We consider the use of the K-means algorithm and the improvement of the already existing elements by concatenating the information at cluster level with the best form of matching and similarity we can obtain a detection box with ideal IOU [9].

C. Discussion on the experimental procedure

The algorithm uses the DarkNet-53 network in its process, which is so named because it contains about 53 convolutionary layers. We can say that these convolutionary layers combine exactly three maps with characteristics at different scales, using high resolutions of lower characteristics, semantic information with high standards and characteristics alike. The algorithm network is sampled approximately 32 times when the image enters the processing field, a high factor in both the receptive field due to high characteristics and filtered surface information, sometimes losses are determined after the convolution of the sampled multilayer. That is why it is ideal to use in the conditions of facial detection and objects in images for the way in which the whole process of analysis and processing takes place [10]. For larger images and features, anchor tapes, and predictive analysis of the hold to be analyzed are used using those standardized scales to be able to have an overview even through those delimitation grids [11]. We consider this module developed within the main project an important step in achieving a completely autonomous system dedicated to safe driving, a necessary and useful system for most road users, especially when road accidents are a major cause of death. The situation that can be radically changed in the conditions of using dedicated systems pre-installed in new cars. Thus, according to studies, equipment, and communications such as VLC, V2I, V2V, and V2X can significantly reduce road accidents and loss of life.

Fig. 3: Position of features in eye and mouth region

III. Considerations on the Experimental Testing Setup

The analysis and tested methods for locating the detection area were incremented by a static distribution with the shapes of the models loaded in the data set for algorithm training, although some elements may be deformable and the detection may no longer be eloquent. Objects with similar and complementary shapes in the training set can compare some

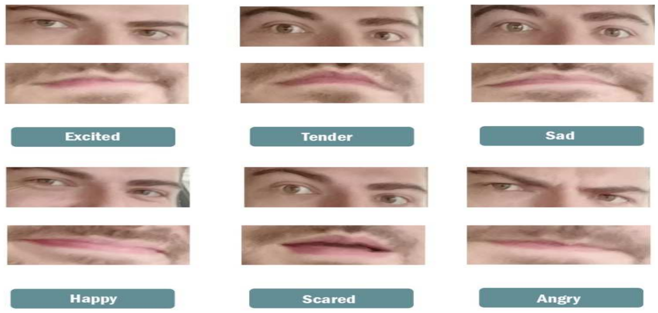

Fig. 4: Models of features with eye and mouth region

even deformable elements by analyzing the eyes and mouth. Changes to the algorithm create an image gradient to describe the characteristics of local images and the contouring of each landmark to remove minimal elements. We can say that the algorithm performs improved extractions, increases the robustness and quality of detection by outlining the areas of the eyes and mouth. Conventional modeling represented by a low-dimensional vector shape analyzes the main comparable eigenvectors by the shapes loaded in the drive models and then the shape obtained by linear transformation is compared with the comparative model.

A. Debate and presentation the models and algorithm

So the method of dividing the face area into several regions is one of the solutions that could lead to a valid result. According to the illustration in Figure 3, we analyze the features for each state presented in Figure 1 with the 6 types of states that can negatively or positively influence the driving style but also the way the driver expresses himself depending on the situation he is in. Thus, Figure 4 shows the models for each previously characterized state that will be found in the analysis and detection algorithm that tries to compare all similar items with the images entered and analyzed in real-time. The created models are based on the elements presented in Figure 4 making a model of information that produce the data set through which it is possible to analyze the predictability and similarities of the characteristics present in the analyzed images. Training models and data analysis according to the created models are analyzed by the following syntaxes:

```
trainm_generator = train_datagen.flow_from_directory(
        trainmodel_data_dir,
        target_size=(48,48),
        batch_size=batch_size,
        color_mode=" redgradient ",
        class_mode='greencategorical')
validation_generator = val_datagen.flow_from_directory(
        validation_data_dir,
        target_size=(48,48),
        batch_size=batch_size,
        color_mode="redgradient",
        class_mode='greencategorical')
filepath = os.path.join("./carsafe_emotion_models/model_v6_{emotiondriver}.hdf
5" )
print(validation_generator.class_model)
{'Excited' , 'Tender', 'Sad', 'Happy', 'Scared' 'Angry'}
```

A classification report is presented in which the criteria of accuracy in Table I, repeal, and information support are introduced. According to the classification report, there is a degree of confusion in the case of emotional states that subsequently led to a low degree of prediction and a non-compliant validation with a totally erroneous answer.

For some data sets that allowed their conversion or smoothing, multiple filters and detections were applied, whether we are talking about images or video frames.

TABLE I. CLASIFICATION REPORT – FROM CONFUSION MATRIX

	Precision	Recall	YV3-score	Total Frame	Time Processing
Excited	0.68	0.69	0.71	127	76ms
Tender	0.17	0.97	0.28	77	102ms
Sad	0.14	0.99	0.24	98	108ms
Happy	0.77	0.89	0.84	67	64ms
Scared	0.55	0.58	0.33	89	92ms
Angry	0.78	0.78	0.71	67	68ms

B. *Experimental evaluation and results*

We can see in the Figure 5 the detected states are: *Excited*, *Happy*, *Scared* and *Angry*, as for the *Sad* and *Tender* the detection was confused by manifesting the states in different ways the facial analysis could not identify the two states because the position of the eyes and mouth is similar and not has elements compared to the introduced models. The developed classical detection systems can sometimes have errors in the evolution of states without an intensive study in neuro-psychic problems. Human emotions and states are different, feelings and manifestations are also different depending on the personality of each. The models created were based on these studies because we must consider each feature as they can influence the degree of accuracy. The procedures for filtering, smoothing, and extracting the facial area was performed by generating and contouring a histogram, then the data set was superimposed on the models created over the components extracted with the facial area so the feature vector could highlight components with similarities really answering or false for each overlapping emotional state.

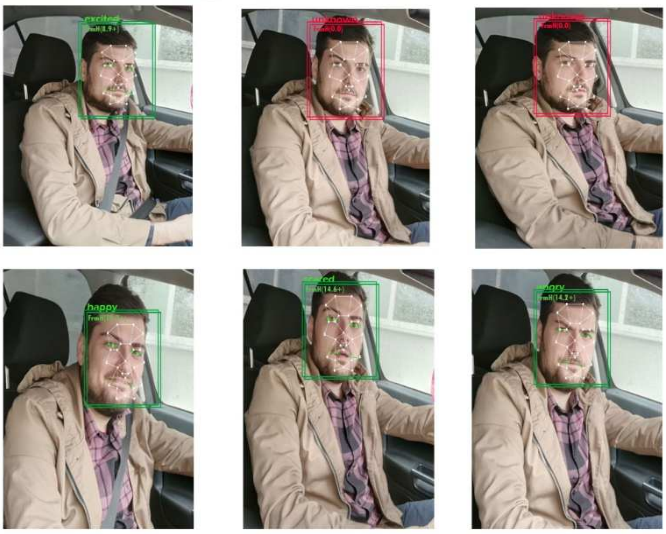

Fig. 5: Detected and undetected emotional states

IV. Conclusion

We can say that the module developed and implemented within a hybrid system of visible light communications offers a research perspective in the field. The analysis and emotional detection of drivers combined with the transmission of information through VLC facilitate the possibility of timely communication with other road users having the ability to react in remedying a dangerous situation. Facial condition determination is present in a platform dedicated to road safety. The models created provide results in certain layers for analysis and detection with low execution times and high processing capacity. The current system is based on RF - VLC communication based on several hardware components that can be adjusted for improved results, including neural processing that would increase system performance leading to simulations much closer to expectations. We want to improve mobility and adaptability depending on low light conditions or non-contrast elements.

We believe that the purpose of the research was achieved by analyzing and determining the main emotional states that negatively influence the driving style, these being used in a traffic safety platform based on hybrid communications.

References

[1] Fact Sheet 358—Road Traffic Injuries, World Health Org., Geneva, Switzerland, May 2016.

[2] P. Ekman and W. V. Friesen, “Constants across cultures in the face and emotion.” Journal of personality and social psychology, vol. 17, no. 2, pp. 124–129, 1971.

[3] E. Zadobrischi and M. Negru, "Pedestrian detection based on TensorFlow YOLOv3 embedded in a portable system adaptable to vehicles," 2020 International Conference on Development and Application Systems (DAS), Suceava, Romania, 2020, pp. 21-26, doi: 10.1109/DAS49615.2020.9108940.

[4] B. T. Nguyen, M. H. Trinh, T. V. Phan and H. D. Nguyen, "An efficient real-time emotion detection using camera and facial landmarks," 2017 Seventh International Conference on Information Science and Technology (ICIST), Da Nang, 2017, pp. 251-255, doi: 10.1109/ICIST.2017.7926765.

[5] M. M. Taghi Zadeh, M. Imani and B. Majidi, "Fast Facial emotion recognition Using Convolutional Neural Networks and Gabor Filters," 2019 5th Conference on Knowledge Based Engineering and Innovation (KBEI), Tehran, Iran, 2019, pp. 577-581, doi: 10.1109/KBEI.2019.8734943.

[6] B. Balasubramanian, P. Diwan, R. Nadar and A. Bhatia, "Analysis of Facial Emotion Recognition," 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2019, pp. 945-949, doi: 10.1109/ICOEI.2019.8862731.

[7] X. Chao and F. Zhiyong, "Facial Expression Recognition and Synthesis on Affective Emotions Composition," 2008 International Seminar on Future BioMedical Information Engineering, Wuhan, Hubei, 2008, pp. 144-147, doi: 10.1109/FBIE.2008.53.

[8] A. Verma, P. Singh and J. S. Rani Alex, "Modified Convolutional Neural Network Architecture Analysis for Facial Emotion Recognition," 2019 International Conference on Systems, Signals and Image Processing (IWSSIP), Osijek, Croatia, 2019, pp. 169-173, doi: 10.1109/IWSSIP.2019.8787215.

[9] K. Ko and K. Sim, "Development of a Facial Emotion Recognition Method Based on Combining AAM with DBN," 2010 International Conference on Cyberworlds, Singapore, 2010, pp. 87-91, doi: 10.1109/CW.2010.65.

[10] V. M. Alvarez, R. Velázquez, S. Gutierrez and J. Enriquez-Zarate, "A Method for Facial Emotion Recognition Based on Interest Points," 2018 International Conference on Research in Intelligent and Computing in Engineering (RICE), San Salvador, 2018, pp. 1-4, doi: 10.1109/RICE.2018.8509055.

[11] Jian-Ming Sun, Xue-Sheng Pei and Shi-Sheng Zhou, "Facial emotion recognition in modern distant education system using SVM," 2008 International Conference on Machine Learning and Cybernetics, Kunming, 2008, pp. 3545-3548, doi: 10.1109/ICMLC.2008.462